Evaluation of *Staphylococcus carnosus* as an Alternative Host Organism for Whole-Cell Biocatalysis

A Thesis Submitted to the University of London for the Degree of **DOCTOR OF PHILOSOPHY**

2014

Pedro Humberto Bixirao Neto Marinho Lebre

THE ADVANCED CENTRED FOR BIOCHEMICAL ENGINEERING

UNIVERSITY COLLEGE LONDON

I, **Pedro Humberto Bixirao Neto Marinho Lebre** confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgments

This project would not be possible without the essential support of my supervisors, Dr. Frank Baganz and Prof. John Ward, at the University College London (UCL). In particular, I would like to thank Prof. John Ward for providing the bench space where most of my research was conducted and the tools required to conduct it. I would also like to give thanks to the lab members in John Ward's group, in particular Dr. Lottie Davis, Dr. Markus Gershater, Dr. Oriana Losito, Dr. Julio Martinez, and Dr. Vishal Sanchania for their technical expertise and moral support. Finally I would like to thank my family and friends for their affectionate support and encouragement throughout the PhD project, without which the weight of many failed experiments would be unbearable.

A special thanks is reserved to the Biotechnology and Biological Sciences Research Council (BBSRC) for providing the funding for this research.

Abstract

The aim of this project is to study the use of Staphylococcus carnosus as an alternative host in whole-cell biocatalysis, using the NADPH-dependent conversion of cyclic ketones by cyclohexanone monooxygenase (CHMO) as the model system.

The use of whole cells in industrial biocatalysis has been actively researched due to their promise as biological factories that can perform complex reactions in environments where the use of isolated enzymes would not be feasible. However, the majority of whole-cell processes rely on a small range of organisms that, while been extensively characterized and easily engineering to express enzyme biocatalysts, have some inherent limitations that hinder their application to many industrial processes. Thus. There is a need to find novel microorganisms that can fill in the shortcoming of those conventional strains.

In this project, a flexible shuttle vector was constructed to allow for the cloning and expression of CHMO in both *E. coli* and *S. carnosus* by using a dual promoter system in which one of the promoters could be replaced. Higher levels of CHMO expression were achieved in *E. coli* strains cloned with the shuttle vector compared to an expression vector used in previous studies. No biocatalyst expression was detected in *S. carnosus*. The shuttle vector was subsequently modified to allow for the quantitative characterization of different synthetic promoters based on the gemone of S. carnosus. A novel *in-silico* promoter selection methodology based on the codon bias was developped to allow for the rational selection of potential genomics promoters. Two synthetic promoters based on sequences upstream of ribosomal proteins rplK and rplJ were subsequently shown to allow for protein expression in both *E. coli* and *S. carnosus.* The stronger of these two promoter was inserted into the CHMO expression vector, but the resulting construct did not manage to express detectable levels of the biocatalyst in *S. carnosus*. It was thus concluded that *S. carnosus* was not a suitable host for biocatalysis with CHMO, and further studies with other enzymes would have to be conducted to access its suitability as a general biocatalytic host.

Table of Contents:

Table of Figures:

Chapter 1: Introduction to the project

1.1- Biocatalysis as an Emergent Technology

Biocatalysis as a field revolves around the idea of using biological systems to perform chemical reactions, either through the use of isolated enzymes or microorganisms. By necessity, the biocatalytic approach relies on our understanding of the biological systems that make up organic life, and subsequently allows us to take inspiration from these systems in the design and application of biocatalysis.

The idea of biocatalysis as a distinct field in which biological entities are purposely engineered for use in human-constructed processes is fairly recent when compared to the much longer history of the subconscious use of biotransformations in processes that permeate our lives. The very familiar process of beer production is one of the oldest examples of an accidental biotransformation by using microorganisms in the production of beverages. The oldest records date from 6000 BC in the “Fertile Crescent” when Sumerians and Babylonians were practicing the art of brewery. Another example of these early biocatalysts is the use of yeast in bread baking by the Egyptians (Liese et al. 2008). These examples illustrate the crucial role that biocatalysis fulfilled throughout history in shaping what we produce and how we produce it.

It was only in the second half of the 19th century, with the discovery of the enantioselective fermentation of tartaric acid by Louis Pasteur in 1857 (Gal 2008), that biocatalysis became an active field of research. Louis Pasteur’s discovery not only shed light on the use of microbial systems for the production of chemical compounds, but also highlighted how biological systems introduced chirality into a chemical reaction. This promoted a surge of research into biotransformations that

culminated in Fischer giving the first description of enzymes as biological machines capable of performing reactions in a stereo-selective manner, by means of the very well known lock-and-key metaphor (Reggelin 2005; Bornscheuer & Buchholz 2005).

The lock-and-key metaphor proved to be an apt description of the mechanisms underlying the activity of enzymes. But it wasn't until Eduard Buchner published a series of studies on alcohol fermentation in the 1870s that enzymes as entities that could perform chemical reactions in the absence of living cells was widely considered (Bornscheuer & Buchholz 2005). Numerous studies since have described how enzymes select compounds on the basis of their stereochemistry and efficiently produce compounds of high optical purity (Thomas et al. 2002; Wohlgemuth 2010).

Thus, as the focus of scientific research shifted to enzyme biochemistry, and the knowledge of an ever-growing repertoire of enzyme-mediated reactions deepened, the notion of using biological systems as machines for chemical synthesis started to permeate a broad range of different research and industrial fields (Illanes 2008; Bornscheuer & Buchholz 2005).

The introduction of biocatalysts opened up a new golden age of discovery and technological improvements. More efficient processes for the production of commercial compounds were instituted, while novel biochemical processes were created by exploiting a ever-increasing catalogue of enzyme catalysts. (Otten et al. 2010; Thomas et al. 2002; Pollard & Woodley 2007; Coward-Kelly & Chen 2007).

It is perhaps apt to draw a parallel between the current state of biocatalysis and that moment in history when Louis Pasteur first described the fermentation of tartaric acid. As his discoveries changed the state of scientific knowledge and challenged the *status quo* at the time, so now is the field of biocatalysis pushing the boundaries of current science, by changing the landscape in with we create products and commodities, and by presenting new and exciting ways of expanding our biotechnological capabilities.

1.1.1- Biocatalysis in the Chemical industry

The chemical industry is an area with a strong impact in a broad range of commercial products used in everyday life, including pharmaceuticals, fragrances, detergents, textiles and biofuels (Sanchez & Demain 2010). The importance of this industry has led to increasing demand for long-term sustainable processes and the production of more specialized products, both of which are difficult to achieve using the tradition chemical approach. While chemical processes can achieve very high yields, they often require high-energy inputs, the use of potentially harmful solvents, and result in the production of large quantities of by-products, which require efficient separation (Thomas et al. 2002).

In this context, biocatalysis has become a highly attractive alternative to the traditional approach. Many enzymes are flexible catalysts that can accept a wide range of substrates, and are able to produce optically active compounds with high enantio- and regio-selectivity in a rapid fashion. This solves several problems of the conventional chemical industry. Firstly, biocatalysis allows for the use of renewable and cheap feedstocks and does not require the numerous protection of vulnerable groups and their subsequent de-protection steps of chemical processes In addition, the highly selective reactions involved in biocatalysis bypass the need to develop complex downstream processes for the separation of unwanted bi-products. Biocatalysis also provides answer to the more recent and increasingly important concern of creating “greener” industrial processes (Woodley et al. 2008).

An example of the relevance of the stereo- and regio- selectivity of enzyme biocatalysts for the chemical industry is the production of pharmaceutical compounds such as neuraminic acid, a precursor to the anti-influenza drug Zanamivir, which is produced by N-acetyl-D-neuraminic acid aldolase (Liese & Villela Filho 1999). The chiral complexity of the substrates and products involved in this reaction would require several protection and de-protection steps during chemical synthesis, while the aldolase can carry out the reaction in a single step without protection. Another example of the advantages of biocatalysts is the use of peroxidases for production of phenolic resins, which are used in numerous industries as molding materials, circuit component s, coatings and adhesives, among others (A. Schmid et al. 2001). The chemical synthesis of these resins

involves the condensation of formaldehyde with several phenol-type, but the toxicity of the former compound as led to a shift of the process into the direct enzymatic polymerization of phenols (Dordick et al. 1987). In this case, the peroxidase allows for conversion to occur in milder condition than the chemical counterparts by using hydrogen peroxide as the oxidant, thereby removing the toxic pollutant from the process waste.

In 2011, it was estimated that the overall industrial biocatalytic market accounted for more than 500 products, excluding diagnostics and pharmaceuticals, with the bulk of the biocatalytic processes being focused on detergent production and food processing (Sanchez & Demain 2010). Several enzymes are currently routinely used as additives in detergents for stain removal, the most common being the serine proteases from *Bacillus* species due to their stability under high temperatures and alkaline pH (Maurer 2004). More recently, second-generation enzymes such as amylases, cellulases and mannanases that are better for the removal of specific stains have been developed (Kirk et al. 2002). Detergent production remains the largest application of enzyme catalysts in chemical industry due to the commercial scale of detergent use (Figure 1).

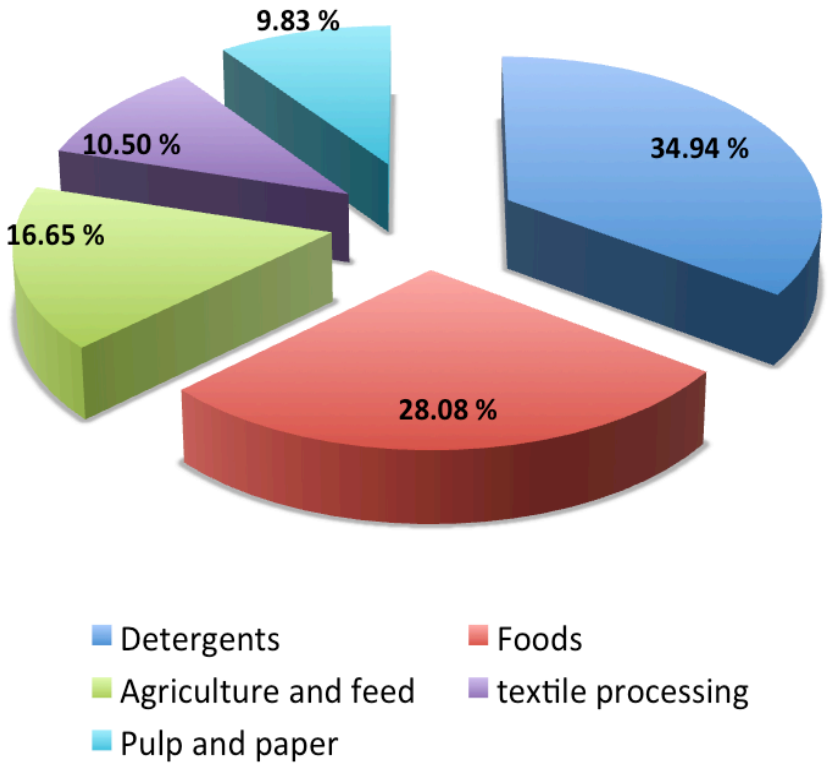


Figure 1- Enzyme market distribution for the Chemical Industry in 2003 (Sanchez & Demain 2010).

The application of biocatalysts to the food industry is more diverse, ranging from flavor production to food preservation. As previously mentioned, microbial biocatalysts have historically played a very crucial role in beer brewing and bread baking. With the development of technologies for enzyme screening and purification, several other food processes have since been put in the charge of biocatalysts. For instance, the production of the low-calorie sweetener Aspartame, which is applied to a broad range of products from soft drinks to ready-made meals, and is composed of the two natural amino acids L-aspartic acid and L-phenylalanine. is catalyzed by the thermophilic zinc protease thermolysin (Bommarius & Riebel 2005). This enzyme not only has an optimal temperature of 50- 60 ^{o}C, but it also allows for a tight control over the regio- and enantio-selectivity of the reaction. Another compound that is used as a sweetener, D-fructose, is converted from D-glucose by a xylose isomerase that is active at 70 ^{o}C in packed bed reactors (Schoemaker et al. 2003). Microbial and fungal proteases

are also involved in food processing as substitutes for calf rennet for the milk clotting process in cheese production (Shieh et al. 2009; Vishwanatha et al. 2010). The use of engineered enzymes as substitutes for the coagulating mammalian proteins chymosin and pepsin in the calf rennet improves the cheese manufacturing process in two major ways: it solves the bottle-neck of limited enzyme supplies found in the calf stomach extracts, as well as preventing heath risks related to bovine spongiform encephalopathy or scrapie (Jacob et al. 2011). Finally, Gram-positive bacteria such as *Staphylococcus carnosus* are used in the curing of raw meat for the preservation of colour, flavor production, and prevention of spoilage by pathogenic organisms (Gøtterup et al. 2008; Simonová et al. 2006).

Several other industrial areas employ biocatalysts to a lesser extent. In textile industries the scouring step of cotton processing, which revolves around the removal of cell-wall components on cellulose fibers, conventionally involves high temperatures and alkaline conditions. This can be avoided by using microbial and fungal pectinases, which allow for the process to be performed in a single step under milder process conditions (Tzanov et al. 2001). In the production of animal feed-stocks, fungal, bacterial and yeast phytases are used to release phosphorus from phytates found in plant-derived feed (Haefner et al. 2005), which cannot be digested by monogastric animals . Since phytates account for 60 to 80% of phosphorous in plant-derived feedstock, the use of enzymes that complement for deficiencies in mammalian digestive systems is essential in order to provide a sufficient diet in the food. In addition, the present of phytates in the human diet has been related to negative heath effects. As a anionic compound, phytate, will bind to several positively charged mineral ions and proteins, thus affecting the uptake of essential dietary minerals like zinc and calcium as well as the activity of proteins they bind to (Greiner & Konietzny 2006). Therefore, there is also a health-related interest in the use of phytases in animal feedstocks and food industries.

Despite the clear advantages of biocatalysis, there are still problems that prevent the application of several promising biocatalysts in large-scale industrial processes. One of the problems is that a number of chemicals and polymers produced in chemical processes are not soluble in aqueous solutions and require reactions to be performed in organic solvents. While several enzymes have been found to be sufficiently stable and show activity in hydrophobic environments, there

is a big proportion of proteins that are not active in such environments (Schmid et al. 2001). Another bottleneck is the struggle of biocatalytic reactions to achieve the high yields of traditional chemical processes, either due to enzyme instability or the need for co-factor regeneration. The successful transition of biocatalysts into the chemical industry is dependent on how cost-competitive the biocatalytic processes are when compared to the more conventional chemical counterparts (Sanchez & Demain 2010). Thus, biocatalytic processes that perform bellow industrial standards are handicapped, as they often require a lengthy optimization process. This in itself becomes a problem, as many companies are unwilling to invest high sums of money into novel processes that require specialized training and the extensive development and optimization of processes prior to application (Thomas et al. 2002). On the other hand, the advent of the genomic era has opened a window to the discovery of novel biocatalysts that can operate in more extreme environments, while also sponsoring developments in genetic engineering techniques that allowed for the optimization of current biocatalysts. These technological advances will be discussed in more detail in a later section.

1.1.2- Biocatalysis in the Medical Sciences

Biocatalysis also plays a very important role in the development of new-generation therapeutics. As the bank of natural compounds available for therapeutic use dwindles and becomes less effective against a growing number of pathologies, the pharmaceutical industry has led the search for novel small molecules that are either based on the natural counterparts or purely novel synthetic compounds (Newman & Cragg 2012). Enzymes, with their tightly defined regio- and enantio-selectivity, are ideal catalysts for the construction of complex structures with several chiral centers and active groups.

In fact, the relevance of biocatalysis to drug discovery cannot be better exemplified than by one of the great medical developments in the 20th century, the discovery of penicillin. This discovery was made by Alexander Fleming, who reported in 1929 the anti-bacterial effects of a substance secreted by a *Penicillium* mold on several pathogenic bacterial cultures (Fleming 1979), later named

penicillin. The fermentation studies on *Penicillium notatum* conducted by Florey and colleagues at both the Sir William Dunn School of Pathology at Oxford university and the Northern Regional Research Laboratory of the US Department of Agriculture constituted the first instance of an antibiotic produced in large scale through the action of a fungal micro-organism (Kardos & Demain 2011). The realization that microorganisms could be used for the development of anti-bacterial drugs highlighted the potential of biological systems as sources for drug discovery, and opened up a new era for therapeutic-directed biocatalysts.

Since the major breakthrough in the discovery of penicillin, biocatalyst systems have been recognized as sources of natural products for therapeutic uses. Thus, several bacterial and fungal microorganisms have been employed in the pharmaceutical industry due to their native ability to produce bioactive secondary metabolites that exert a major impact on several diseases and medical conditions. One example is the anti-cancer antibiotic daunorubicin, which is produced naturally by *Streptomyces peucetius* (Demain & Sanchez 2009). This anthracycline acts by interlacing DNA strands, which prevents transcription, and triggering cell-death through Topoisomerase II DNA digestion. It has been found to be very effective against several types of cancer, including leukemia, lymphoma, and lung cancer. Another important group of natural metabolites was found in the fermentation broth of *Streptomyces avermitilis* (Campbell 2012). These compounds, named avermectins, are 16-membered macrocyclic ketones that have a strong toxic activity towards parasitical nematodes, but are non-toxic to most other organisms (Dent et al. 1997).

Biocatalysts have also been extensively used in pathways for the synthesis of synthetic compounds. In fact, biocatalysis has been the major driving force behind the expansion of the repertoire of bioactive compounds that replace the dwindling effective natural products available to drug manufacture. The production of the anti-hypertensive drug Omapatrilat is a clear example of how biocatalysts are integrated into drug synthesis. This compound is a complex chiral molecule that can be used to treat hypertension and congestive heart failure by inhibiting both the angiotensin-converting enzyme and an endopeptidase involved in the stimulus of the production of cGMP and consequent vaso-dilatation (Seymour et al. 1991) by blocking the activation of the cGMP cyclase. One intermediate in this

drug symnthesis, (S)- 6- hydroxynorleucine, is synthesized through the activity of four enzymes. First a D amino acid oxidase and catalase are used to make 2-keto-6hydroxyhexanoic acid which in turn is acted upon by glutamate dehydrogenase making (S)-6-hydroxynorleucine. The reaction uses ammonia and the coenzyme is NADH. The NADH is regenerated by using glucose and a glucose dehydrogease from *Bacillus megaterium.* (Patel et al. 2003). The use of these enzymes solves some of the problems of chirality and multiple sensitive functional groups present in Omapatrilat. Perhaps a more relevant example of high-impact biocatalysis of synthetic drugs is the production of an intermediate to the neuraminidase inhibitor Tamiflu® from Roche. This intermediate compound, shikimic acid, is also an intermediate of aromatic amino acid synthesis, and through pathway engineering strains of *E. coli* were created to allow for the bulk production of this compound (Panke & Wubbolts 2005; Escalante et al. 2010).

One class of enzyme catalysts that has been the focus of intensive research in the context of therapeutics is cytochrome P450s (CYP) (Figure 2). These enzymes are *haeme*-containing monooxygenases that catalyze a variety of different oxidation reactions on a broad range of substrates, including active secondary metabolites of human metabolism, such as hormones and signaling molecules, as well as intermediates of drug metabolism (Urlacher & Girhard 2012). Thus, the pharmaceutical relevance of these enzymes stems not only from the fact that they are able to catalyze a large repertoire of processes, but also because they interact directly with bioactive compounds such as steroids and xenobiotics. Correspondingly, a number of drug synthesis processes using microbial p450s have been established. For instance, the production of Pravastanin, (an HMG-CoA reductase inhibitor that suppresses cholesterol biosynthesis), from compactin by a *Streptomyces sp.* cytochrome P450 has been developed into a lucrative commercial process (Urlacher & Eiben 2006).

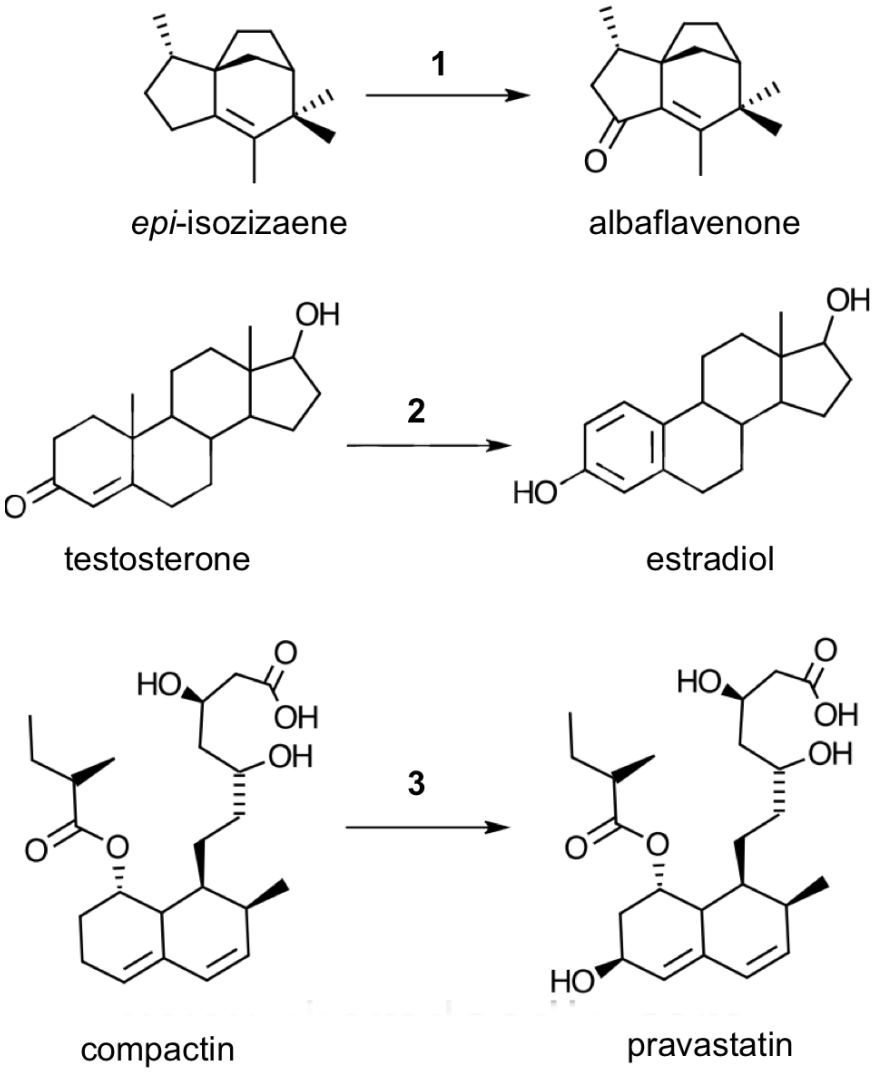


Figure 2- Industrially relevant oxidations and hydroxylations conducted by cytochrome P450s from different native organisms: CYP170A1 from *Streptomyces coelicolor* A3(2) converts *epi*-isozizaene into the antibiotic albaflavenone (reaction 1) (Urlacher & Girhard 2012) ; CYP19, which is abundant in the human liver, catalysises the conversion of testosterone into estradiol (reaction 2), which can be subsequently converted into anti-cancer compounds (Tsuchiya et al. 2005); the anti-cholesterol drug pravastatin can be economically produced using a CYP from the soil bacteria Streptomyces sp. Y-110 (reaction 3) (Park et al. 2003).

Human CYPs have also been extensively researched as potential therapeutic biocatalysts due to the fact that they are native to the end users of the pharmaceutical industry. Several pharmaceutical companies have successfully cloned and expressed human CYPs in microbial and yeast systems for large-scale synthesis of bioactive metabolites. One example is Novartis, which tested 14 human CYPs from *E. coli* systems for synthesis of bioactive metabolites from more than 60 different compounds (Schroer et al. 2010). The results highlight the versatility of this class of enzymes, as several of them were able to synthesize metabolite building blocks that could be later used in drug production. More recent therapeutic strategies focus on the use of CYP as prodrug activators for treatment of cancer cells, thus using the target cells as the centers of biocatalytic focus. For instance, the human CYP2B6 was delivered to cancer cell culture *via* a retroviral vector and was shown to successfully activate the prodrug cyclophosphamide (CPA) into a more active compound that damaged DNA in the cell, thus inhibiting tumor growth (Kan et al. 2001). This approach to chemotherapy offers several advantages to the conventional methods, as it allows for the specific targeting of the toxicity of drugs to cells that express the recombinant CYP, where they are activated. Additionally, human recombinant CYP expression is not in itself toxic to cells, and will not trigger a native immune response.

The examples given above are just a sub-section of the pharmaceutical processes that hinge on enzyme catalysts or whole-cell systems. The heavy investment of companies and academic research has promoted a bloom in the number as well as the types of biocatalysts, which in turn paved the way to the production of more complex bioactive metabolites and drug compounds. Correspondingly, the impact on the pharmaceutical industry on healthcare is partially dependent on breakthroughs made in the field of biocatalysis. Recently, several approaches have been adopted not only to find several novel biocatalysts of medical importance, but also to change the nature of biocatalysts so that they can perform a broader range chemical reaction, and subsequently grant access to a wider variety of drug compounds. These approaches will be discussed in the next section.

1.2- Enzyme versus Whole-Cell Biocatalytic Systems:

In the examples mentioned above, both isolated enzyme system and whole-cell systems were discussed in equal measure. However, there is a clear distinction between these systems that cannot be ignored, as they do not stand on equal footing for every chemical process that would require a biocatalyst. Therefore it is worth spending some time analyzing both systems in parallel and compare their advantages and disadvantages.

Enzymes are powerful biological machines, optimized to catalyze a wide variety of reactions in a enantio- and stereo- selective manner, while producing optically pure products in a single reaction that does not require large inputs of energy and hazardous chemical compounds (Bommarius & Riebel 2005; Thomas et al. 2002). Thus, the use of these machines in-vitro *in vitro* offers a very attractive alternative to conventional organic chemistry, which normally relies on several steps and cannot provide pure products in a very straightforward manner. A very relevant example in the context of this PhD project is the use of enzyme biocatalyst for Baeyer-Villiger oxidations in industry. As it will be discussed in more detail later in this chapter, Baeyer-Villiger oxidations refer to reactions in which an oxygen atom is introduced into the structure of ketones near the carbonyl groups, resulting in the production of esters, or lactones in the case of cyclic ketones (Mihovilovic et al. 2002). In the chemical process, peracids or hydrogen peroxide are used as the reagent for the reaction, but the use of these compounds poses a major challenge for industrial scale Baeyer-Villiger oxidations due to their reactivity and toxicity (Kamerbeek et al. 2003). By contrast, the use of enzyme catalysts not only simplifies the reaction process by removing the need for protection and de-protection steps to prevent by-product formation with the use of reagents, but also presents a greener alternative to the use of large volumes of hazardous compounds.

In many chemical processes, isolated enzymes are seen as the preferred biocatalysts over whole-cell systems. One of the major advantages of isolated enzyme systems is the low degree of system complexity, when compared with whole-cell systems with secondary metabolic pathways and several potential mass-transfer barriers. In this respect, working with enzyme biocatalysts leads to a simplification of design for any catalysis process (de Carvalho 2011). Another related advantage of isolated-enzyme systems is the turnover of products, which are not dependent on mass-transfer limitations across the cellular membrane, and therefore often result in higher yields (Burton et al. 2002).

The major downside of isolated-enzyme system is the stability of the biocatalysts that operate under stringent microenvironments. Devoid of the homeostatic environment of the intracellular space, enzymes become highly susceptible to small shifts in reaction conditions such as temperature, pH, and non-conventional solvents (Illanes et al. 2012). Correspondingly, several strategies have been developed in the attempt to solve this problem. One of these strategies that has been greatly aided by advances in sequencing technology, is the metagenomic approach, whereupon libraries of metagenomes are screened for enzyme biocatalysts that are inherently optimized for a specific process. One such approach is the screening of DNA from extremophile organisms for the identification of biocatalysts such as lipases and proteases that are stable in organic solvents (Gupta & Khare 2009).

Recent developments in the field of genetic engineering have also allowed for synthetic approaches that focus on the optimization of existing biocatalysts instead of searching for novel enzymes. One of these methodologies, directed evolution, relies on the use of site directed mutagenesis or genetic recombination to create large libraries of iterations from an single enzyme catalyst, which are subsequently screened for increased stability, higher activity, or even novel substrate specificity (Goldsmith & Tawfik 2012; de Carvalho 2011). The great advantage of this strategy, and the reason why it is currently one of the preferred methodologies for optimization of a biocatalytic process, is the fact that a large number of active iterations can be created without a comprehensive understanding of how protein sequence and function are co-related. Of course, in many practical

cases, a prior understanding of the enzyme's mechanism of action and structure greatly improve the chances of creating an active library of optimized iterations.

Another strategy largely employed in industry to increase the stability of biocatalysts is enzyme immobilization, either through cross-linking onto a support, or through absorption or entrapment onto a carrier. While there is no universal rule to predict how different biocatalysts will behave after immobilization, in many cases this strategy has been shown to increase the overall structural stability of the biocatalyst, its tolerance to pH and temperature, and to increase activity (Singh et al. 2013). An added advantage of immobilization is the easy downstream separation of the reacting biocatalyst from the finished product.

While the development of several genetic, genomic, and process strategies allowed for the optimization of enzyme biocatalytic systems and even the creation of novel ones, stability is not the only difficulty in working with these systems. There are numerous cases of highly desirable biocatalysts that require co-factors such as NAD(P)H and FAD for efficient catalysis, which are often very expensive compounds to produce and subsequently constitute a large economical cost in a industrial-scale process. Aside from a couple of examples where hybrid biocatalysts were created by coupling an active enzyme with a co-factor regeneration domain (Torres Pazmiño et al. 2009), co-factor regeneration is still a major limitation to the development of economically viable processes for this type of biocatalyst. Many methodologies used to tackle this problem rely on two enzyme systems, whereupon the activity of the primary biocatalyst is complemented with a second protein and substrate that recycles the co-factors. Some recently developed recycling systems involve the use of enzyme catalytic loops such as glucose-6-phosphate dehydrogenase and formate dehydrogenase systems to regenerate NADPH (Mihovilovic et al. 2002). Another closed-loop catalytic system involve the use of an alcohol dehydrogenase that converts the alcohol substrate into a ketone, reducing $NADP^+$ in the process (Willetts 1997). The problem with such two enzyme systems is the need to supply *in situ* conditions that are compatible with both enzymes, which considerably increases the complexity of the system.

Other types of biocatalysts that are not easily adapted to isolated-enzyme systems are membrane-associated proteins, which require the physical contact to

structures analogous to cellular lipid membranes. A more crucial limitation of membrane bound biocatalysts is the challenge of purifying the protein from the membrane without a significant loss of activity, which often requires complex and cost-demanding protocols (Lin & Guidotti 2009).

While enzyme catalysis have warranted much of the attention due to their immediacy as biological machines and the flexibility that they offer, it is often easy to forget that the first steps into the area of biocatalysis were taken by whole-cell systems. Here, we will only briefly discuss the advantages and disadvantages of these systems over isolated biocatalysts, in order to re-iterate the importance of whole-cell biocatalytic systems as viable options for industrial and research biocatalysis. A more in-depth discussion of these systems will follow in the next sections.

Working with whole-cell systems offers a different approach that bypasses crucial hurdles of isolated enzyme biocatalysis. On one hand, cells have natural recycling systems and thus do not require the use of expensive co-factors or secondary protein catalytic systems to replenish co-factors. On the other hand, use of host cells avoids the need for protein purification and problems with protein instability, as the biocatalysts are expressed and maintained in a tightly controlled intracellular micro-environment that is physically protected from external changes (Kuhn et al. 2010). As mentioned above, these factors represent major difficulties for the development of cost-efficient biocatalytic processes, particularly on the industrial scale. Whole-cell systems are also self-recycling, by which it is understood that the expressed biocatalysts are in a constant cycle of renewal that is concomitant with the metabolic activity of the host. It follows that the biocatalyst recycling step in a whole-cell biocatalytic process is easily achieved by re-growing the host cells (Pollard & Woodley 2007). Therefore, there is an argument to the lower cost of using whole-cell systems as direct catalysts when compared to the costs of protein purification, the R&D needed for stability optimization, and the recycling strategies often associated with isolated-biocatalyst systems.

However, whole-cell systems also have major drawbacks, particularly the added complexity of whole cells compared to the isolated enzyme systems in terms of the number of factors the affect the yields of a given biocatalytic process. For instance, the presence of several metabolites as well as the desired products

in the fermentation broth requires more extensive downstream processes to extract the products to the desired degree of purity. In addition to increasing the number of undesired compounds in the output of the biocatalytic process, some cellular metabolic pathways might directly compete with the desired biocatalytic reactions, resulting in a decreased yield through the degradation of the products or further conversions of the latter into undesirable by-products. One such example is the use of baker's yeast strains to perform Baeyer-Villiger oxidations with cyclohexanone monooxygenase from *Acinetobacter* sp. NCIB 9871. While the biocatalytic conversions were mostly successful, the presence of native proteins that competed with the Baeyer-Villiger oxidation by reducing the ketone substrates ultimately argued against the use of baker's yeast as suitable hosts (Stewart et al. 1996; Kayser 2009).

Another crucial limitation of whole-cell biocatalytic processes is that in several cases these systems struggle to achieve product yield that could be deemed acceptable for industrial processes (Pollard & Woodley 2007). The problem here is two fold. On one hand, the ability of the lipid cell wall to maintain the homeostatic nature of the intracellular microenvironment works as a permeability barrier against the diffusion of certain biocatalytic substrates that are either too hydrophobic, hydrophilic, or large to cross from the extracellular broth to the space where biocatalysis occurs. Recent attempts to counter-act this limitation relied upon genetic manipulation of cell wall synthetic pathways (Ni & Chen 2004), which manage to increase the permeability of the membrane to large hydrophilic compounds. On the other hand, some substrates and products are inherently toxic to cells, and therefore cannot be present in the biocatalytic environment above the toxicity threshold, which in many cases results in very low overall yields. Consequently, in these biocatalytic processes strategies are employed to regulate the concentration of the toxic compounds throughout the reaction, either through establishing a chemostat, where substrate and product are continually added and removed, or through the use of *in situ* resins with high affinities to the toxic compounds (Lye & Woodley 1999; Woodley et al. 2008).

It is clear from this brief analysis of the two type of biocatalytic systems that there is no straightforward answer as to which one is the best. The choice between isolated enzymes and whole-cell biocatalysts will largely depend on the type of

biocatalytic process, its operation conditions and the biocatalysts' requirements. In processes where reaction conditions are mild and a fast turnover is required the use of isolated-biocatalysts might be favored, whereas in situations where the biocatalysts are not stable or require co-factor regeneration the use of whole-cell systems might be more economically viable. For the sake of this PhD project, the subsequent sections of this chapter will focus primarily on the discussion of whole-cell systems. However, we do not wish to prioritize the later over isolated-enzyme systems, but rather show that whole-cell biocatalysis is a viable option in the context of this project's topic.

1.2.1- Conventional Hosts employed in Biotechnology

The current range of microorganisms used for biotechnology is as large as it is wide reaching, aided as it has been by advances in Genetic Engineering techniques as well as the genomic foot printing of several species, which has allowed for an easier characterization and manipulation of such species. Numerous gram-negative and gram-positive bacteria, as well as fungi and mammalian cells are routinely employed as hosts for pharmaceuticals production, as components in biofuel refineries, as preservers in ripening agents in food industry, among others roles.

However, despite this diversity, the gram-negative bacteria *Escherichia coli* can be still singled-out as the most widely used host in biotechnological processes, for both historical and practical reasons. Historically, *E. coli* is the best-characterized organism other than humans. Having been discovered and cultured in 1855 by the German bacteriologist Escherich in the fauna of the human colon*, E. coli* was routinely used as a model for biochemical studies before any other strain (Waites et al. 2009). It was also the first organism to produce human insulin as a genetically engineered therapeutic agent approved for human use, in 1982 (Glazer & Nikaido 2007). In practical terms, *E. coli* strains are organisms that can grow rapidly under temperate conditions, with a doubling time of 30 to 20 minutes, and achieve high biomass yields using simple and cheap carbon sources like glucose (Yang et al. 2011). In addition, as the production of human insulin showed, *E. coli*

strains have a great versatility in the nature of proteins they can stably express. This capacity, together with the vast repertoire of genetic systems developed for heterologous protein expression and the ease of genomic engineering, have made *E. coli* into the staple host system in many industrial processes (Lee 2009).

One such process is the production of amino acids, which are routinely used in human nutrition and consequently are in high demand. The use of bacteria as biocatalytic factories for the production of these compounds is preferable to the use of chemical processes, not only because amino acids are natural by-products of metabolic pathways in many bacterial strains, but also these strains permit the use of cheaper and milder reaction processes. Genetically engineering *E. coli* strains, in which specific metabolic genes have been knocked-out or up-regulated, are routinely used for the production of individual amino acids as more economically viable alternatives to chemical synthesis. For instance, the industrial production of L-threonine is dependent on the use of genetically modified *E. coli* strains in which the pathways for the production of this amino acid are up-regulated, and consequently can achieve yields of 100 g/L (Breuer et al. 2004). Another example of a mainstream industry in which *E. coli* is actively used is the production of Inosine-5'-phosphate (5'-IMP), which is a main flavor enhancer. In this case, a *E. coli* strain expressing an acid phosphatase/phosphotransferase gene from *Escherichia blattae* is used to achieve yields of 156 g/L of 5'-IMP (Ishige et al. 2005), making it the most efficient and economically viable means of producting this compound. Other examples range from the production of chiral alcohols as single drugs to the production of novel carotenoids through pathway engineering (Ishige et al. 2005; Beloqui et al. 2008).

However, *E. coli* is not a perfect industrial workhorse, and the realization of its limitations has been a major driving force in the search for alternative hosts. Genomic engineering has been very useful in partially mitigating many problems of *E. coli* –mediated processes, but some of the limitations cannot be easily bypassed through genetics. These include the inability to perform post-translational glycosylation, which is required for many mammalian proteins (Lee 2009), and inability to survive in extreme conditions, such as in organic solvent-based medium or at extreme temperatures. For instance, processes for the the production of biofuels like ethanol often involve the use of high temperatures, which facilitate

extraction of the product, and high concentrations of solvent, both conditions excluding the use of most *E. coli* strains in preference to extremophiles like *Thermoanaerobacterium saccharolyticum* (Shaw 2008). In other biocatalytic processes, as in Baeyer-Villiger oxidations of cyclic ketones (see section 1.6), the substrates and products of the process are toxic to *E. coli* strains due to accumulation in the cell membrane.

Many of the limitations mentioned above have been bypassed by employing another equally successful group of microorganisms, collectively designated as yeasts. These predominantly unicellular eukaryotic fungi are perhaps the oldest organisms use in the production of human commodities such as bread and beverages, and therefore have advantage over *E. coli* of having been adapted for human use over millennia, (Waites et al. 2009). In addition, yeast like *Saccharomyces cerevisiae* are as well characterized and easily manipulated the *E. coli* counterparts, with an equal kit of tools for genetic engineering and protein expression being available commercially. They also have the added advantage of being more closely related to human and other mammalian cells in genetic terms, and therefore are able to express proteins that are not stably expressed in their E. coli counterparts. Thus, yeasts like *S. cerevisiae* and *Pichia pastoris* are major contributors to key industries such as the production of alcoholic beverages, fuel ethanol, and human proteins, including antibodies and insulin (Waites et al. 2009).

However, despite their advantages, neither *E. coli* nor yeast strains are able to cover the myriad of biocatalytic processes that would benefit from the use of microbial hosts. Thus, as the biomedical and chemical industries shift to a pro-biocatalyst age, there is a ever-pressing impetus to expand the repertoire of microorganisms that can be efficiently adapted and optimized to existing industries, as well as provide the backdrop to the discovery novel processes.

1.3- *Staphylococcus carnosus* as an Alternative Host for Biocatalysis

The name *Staphylococcus* (greek- *staphyle*: “bunch of grapes”; *coccus*: “grape”) was first used by Sir Alexander Ogston in 1882 to describe a new form of micrococci involved in pyogenic abscesses in humans (Cohen 1972). As the name indicates, this type of bacteria was distinct from other types because it formed clusters (Figure 3).

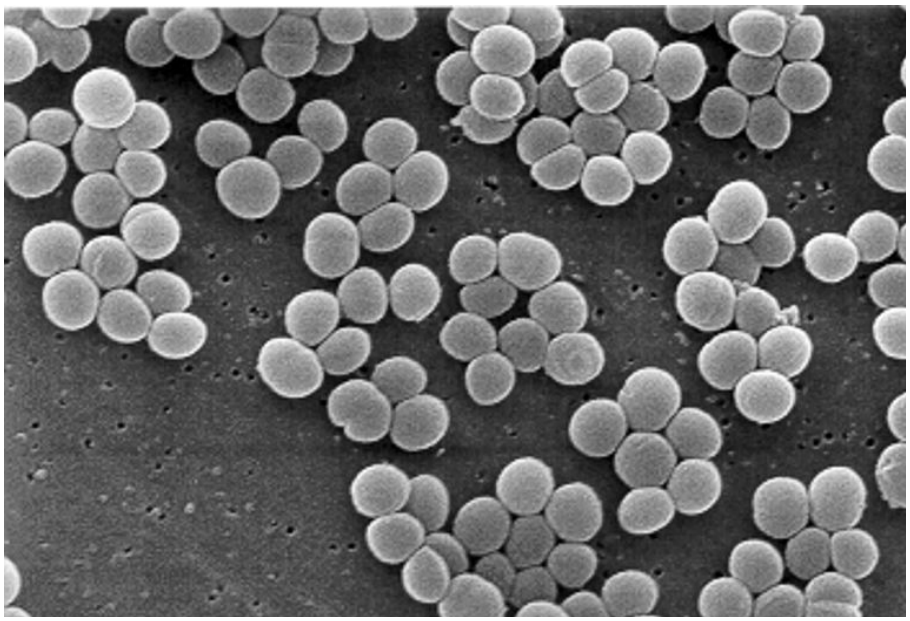


Figure 3- *Staphylococcus aureus* organised into grape-like structures (Reproduced from the Centers for Disease Control and Prevention’s Public Health image Library, image #6486)

).

The definition proposed by Ogston for the genus *Staphylococcus* is still used today, albeit complemented by biochemical, genomic, and proteomic analysis. In the 1920s staphylococci species were re-classified as part of the genus *Micrococcus* under the argument that there were insufficient distinctions between the two species. Only in 1955 was the genus *Staphylococcus* separated from other

cluster-forming cocci on the basis of growth and production of lactic acid from glucose under anaerobic conditions.

Genetic data came to add another criterion to the separation, as staphylococci have lower GC genome content (30-40%) than the other genus. Recently, due to the development of new technologies for DNA and biochemical analysis, together with increasing database for genome sequences, a more extensive characterization of the genus *Staphylococcus* has been achieved, and some species that were previously considered under other genus are now been re-classified as staphylococci (Schleifer & Fischer 1982; Suzuki et al. 2012).

The species *Staphylococcus carnosus* was first isolated from meat fermentation products, hence the name, and was subsequently characterized incorrectly as part of the genus *Micrococcus*, a classification based on phenotypic traits. DNA analysis and DNA fingerprinting techniques came to play a crucial role in the re-classification of the species by dividing the family *Macrococcaceae* into two genetically distinct genera; *Micrococcus* and *Staphylococcus.* These techniques also led to the characterization of S. *carnosus* as a separate species from other coagulase-negative staphylococci found also in fermented meat, such as *Staphylococcus piscifermentas*, and *Staphylococcus xylosus* (Pantůcek et al. 1999).

Schleifer and Fischer (1982) describe this novel species as a 1 to 3 mm diameter cocci colonies that can occur as a single or in pairs, Gram-positive as classified by Gram staining, and not showing motility or spore forming ability. They also characterized S. *carnosus* as a facultative anaerobe, producing equal amounts of D- and L- lactate from glucose during anaerobic growth.

As a Gram-positive bacterium, the intracellular space of *S. carnosus* is enclosed in a single inwards facing lipid membrane and outwards facing peptidoglycan wall that confers the physical robustness of Gram-positive bacteria. The peptidoglycan wall is a defining structure of Gram-positive bacteria, as it is substantially thicker (20-80 nm) than Gram-negative counterparts (7-8nm), and acts as the initial permeability and physical barrier between the extracellular and intracellular environments, as opposed to Gram-negative bacteria where it lies between the two lipid membranes (vollmer et al. 2008). *S. carnosus* exhibits

several phenotypic and genotypic characteristics that are considered highly desirable for whole-cell biocatalysis.

1.3.1- The Staphylococcal Cell Wall

The general structure of the peptidoglycan wall is a mesh composed of layers of alternating beta- 1.4 linked N-acetylmuramic acid (NAM) and N-acetylglucosamine acid dimers (NAG). Layers are further cross-linked *via* peptide stems that are connected to the N-acetylmuramic acid residue. In addition, the muramic acid can also serve as an anchor for wall teichoic acids (Xia et al. 2010) (Figure 4).

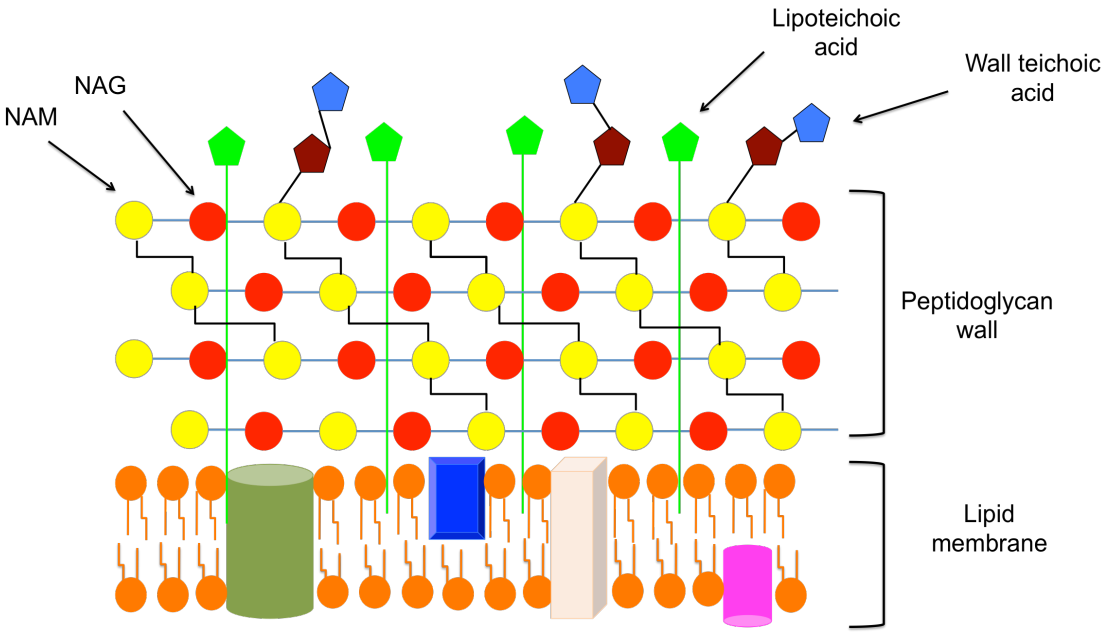


Figure 4- diagram showing composition of gram-positive bacterial cell wall (anon).

1.3.2- Non-pathogenicity

The *Staphylococcus* genus is made up of 36 species, most of which can be found as colonizing bacteria in the skin and respiratory tracks of various higher mammals, including humans. In addition to being inherent to the bacterial fauna present in our bodies, some species, such as *Staphylococcus aureus* and *Staphylococcus epidermis* are aggressive invaders that can cause severe human infections. Of particular note are MRSA and VRSA (methillin resistant- and vancomycin resistant- *Staphylococcus aureus*), which constitute the main causes of skin, pulmonary, and prosthetic-device related infections in hospital patients due to their resistance to antibiotic treatment (Rosenstein & Götz 2010). It is therefore understandable that much of the scientific interest in this genus has been driven by the need to find novel, effective treatments for these diseases.

Staphylococcus carnosus stands apart from most of the other species of the genus, partly because it is not normally found in the bacterial fauna of higher mammals, and partly because it is a non-pathogenic bacterium, despite its phylogenetic proximity with the more pathogenic strains of the genus. The answer to these divergences can be found in the genetic comparison between the different species.

The *S. carnosus* genome is 2.57 Mb long, 86% of which encodes for functional genes. Despite the small size of this genome relative to other staphylococci, S*. carnosus* shares 50 % of its gene products with other pathogenic species like *S. aureus, S. epidermis*, *S. saprophyticus*, and *S. haemolyticus* (Götz et al. 2004). These orthologous genes are organized in core region around the origin of replication that maintains an highly conserved gene structure across the different species. While 25% of the shared gene products do not have a known function, it is assumed that this conserved region encodes mainly for housekeeping genes, which are involved in essential metabolic processes and therefore are required for efficient metabolic functions in the different species.

The point at which *S. carnosus* starts to diverge from its pathogenic cousins is in the absence of toxin genes that characterize the pathogenicity of the other

species. Superantigens like the toxic shock syndrome toxin-1 (TSST1), which is one of the main toxins involved in food poisoning, or the pore-forming Parton-Valentine leukocidin (PVR), do not have any corresponding homologs in the genome of *S. carnosus* (Götz et al. 2004). Additionally, it does not contain any adhesion proteins, such as fribronectin and fibrinogen-binding adhesins, which are essential for tissue invasion and biofilm formation. It is possible that the absence of these proteins explains why *S. carnosus* does not occur as a natural component of bacterial fauna of higher mammals.

However, the latter is not completely devoid of virulence-associated genetic elements. In fact, 41 proteins from a group of 187 *S. aureus* COL proteins associated with pathogenicity and antibiotic resistance are shared with *S. carnosus* (Becker et al. 2007). Since these shared proteins are not pathogenic in *S. carnosus,* it has been hypothesized that their function is more directly related to general regulatory pathways that are also involved in the expression of toxic genes.

Additionally, the genome of *S. carnosus* contains genomic islands that are characterized by a lower GC content and were possibly horizontally transferred from other genomes. These islands contain several genes of unknown function that share a weak homology with pathogenicity-associated genes. For instance, one of these genes codes for putative IgG-binding protein with 30% with protein A from *S. aureus*, but it is clear from experimental studies that *S. carnosus* does not exhibit IgG binding capacity. Similarly, there are also 2 putative exotoxin genes present in these clusters, despite the fact that *S. carnosus* does not excrete toxins (Rosenstein et al. 2009). It is clear that these genes are non-functional as pathogenic markers, either because they exert a different function despite the homologies, or because their expression in *S. carnosus* is disrupted.

Aside from the lack of toxic genes, *S. carnosus* also contains very few mobile elements such as genomic islands or IS elements that are commonly found in the genomes of pathogenic staphylococci, and are associated to the quick adaptability of these species to host's environment or to conferring antibiotic resistance. For instance, an IS element in VRSA (vaconmycin resistant *S. aureus*) designated IS256 has been found to be inserted into the promoter region of a two-

component regulatory system involved in antibiotic resistance. This insertion results in the overexpression of proteins of the regulatory system, probably by disrupting the promoter's repression mechanism, which in turn increases the resistance of VRSA to vancomycin and other antibiotics (Becker et al. 2007). The absence of these IS elements in *S. carnosus* argue for a more stable, albeit less adaptable bacterial system.

1.3.3- Industrial Applications

Historically, *S. carnosus* was first isolated from cured meat products, so it is very likely that this bacterium was being unintentionally used as starter culture in the meat curing process for many years before the industrial use of bacteria in food processing was even acknowledged.

In the traditional curing processes no starter cultures are added, and instead the flavor and aroma are achieved by adding a mixture of salt and sugars. However, economic pressures and the desire to control fermentation conditions have driven industry into the use of starter cultures as a means to enhance product consistency and decrease fermentation times (Masson et al. 1999).

The industrial curing of raw sausages involves the use of lactic acid producing bacteria such as *Lactobacillus* or *Pediococcus*, together with other starter cultures such as *Staphylococcus* and *Micrococcus*. *S. carnosus* is regularly used as a starter culture in Greek, Italian and Turkish sausage production because of its ability to maintain the color of fermented meat, to produce flavor and aroma, and to reduce nitrate into ammonia (Masson et al. 1999; Søndergaard & Stahnke 2002).

The curing process varies from country to country, but it generally involves a pre-inactivation step (aging), a fermentation step, and a drying and storage step. Throughout the several stages, the pH of the sausage is dropped from 6 to between 5 and 4, a decrease that correlates with the increase in lactic acid production by lactic bacteria during the fermentation step. A useful side-effect to the production of lactic acid and subsequent decrease of the environment pH

throughout the curing process is that it works to prevent the settling of pathogenic bacteria. Individual stages in the process also require the use of unique temperatures and humidity, such that temperature and water activity vary greatly throughout the duration of the process. In addition to these varying factors, meat is often cured in environments with high salt, nitrate and nitrite concentrations.

Therefore, as a starter culture, S. *carnosus* has the advantage of being highly adaptable to the various osmotic, temperature, and chemical pressures of the environment in which it coexists. Sondergaard and Stahnke (2002) showed that while there is a positive interaction between temperature and pH on the growth of starter cultures, high salt concentrations seems to have a negative effect on most cultures except for *S.carnosus*, which has optimum growth at 10% w/v NaCl. The resistance to high salt concentration can be explained genetically, as *S. carnosus* contains several systems of osmoprotection, as well as five sodium ion/proton antiporters and homologs to macheno-sensitive ion channels (Rosenstein et al. 2009).

S. *carnosus* is crucial for development of flavor, aroma, and preservation of appearance during the drying stage of the process. Several studies indicate that flavor and aroma emerge from common products of S. *carnosus* catabolism of branched-chain-amino-acids (BCAA) and aromatic amino-acids. In particular, the production of the corresponding methyl-branched aldehydes, acids and esters from the catabolism of leucine seems to have a strong impact on aroma. Two distinct pathways for the production of these compounds have been proposed (Masson et al. 1999; Madsen et al. 2002), and both involve the conversion of leucine into α-keto*iso*caproic acid as the starting point.

Another group of compounds produced by *S. carnosus* in the course of raw-meat curing that seem to have influence on the flavor of meat products are methylketones, which are formed from the β-oxidation of free fatty acids (Talon & Montel 1997; Talon et al. 1998) . This metabolic capacity to produce ketones from fatty acids may serve a double purpose: as well as contributing to the flavour of the cured meat, the reaction leads to the decrease of free fatty acids which might be available for chemical oxidation. Chemical oxidation through interaction of fatty

acids with peroxide species is believed to be the main factor affecting rancidity of meat products (Montel et al. 1998).

In addition to flavor and aroma production, *S. carnosus* plays a crucial role in the development and maintenance of red pigmentation during the curing process. The red pigmentation of cured meat products is maintained through the addition of nitrate into the fermentation mix, which is reduced to the highly reactive nitrite during curing. In turn nitrite is converted to nitric oxide (NO), which reacts with the haeme group of the meat's myoglobin ($MbFe^{II}$), forming the pink-red pigmented structure nitromyoglobin ($MbFe^{II}NO$) (Talon et al. 1999). Nitrite also offers other beneficial side effects to the overall cured product, as it prevents the oxidation of lipids, which is normally associated with rancidity, and prevents the colonization of pathogenic bacteria. *S. carnosus* naturally reduces nitrate into nitrite under anaerobic conditions as part of the fermentation process, in which nitrate is used as the electron acceptor during nitrate respiration. Two enzymes that are sequentially expressed under low oxygen conditions, nitrate reductase and nitrite reductase, are responsible for this reaction, as well as the subsequent reduction of nitrite into ammonia (Gøtterup et al. 2008). While *S. carnosus* cannot convert nitrite into nitric oxide, the reduction of former into ammonia only occurs once the nitrate concentration in solution is exhausted. Therefore, nitric oxide conversion can happen concurrently to the reduction of nitrate, by using either reducing agents like ascorbate, or other starter cultures that can perform the reaction (Hammes et al. 2012). By naturally reducing nitrate/nitrite into ammonia, *S. carnosus* also works as regulatory agent that controls the concentration of these compounds in the cured product, and therefore prevents the potential hazardous affects of nitrate/nitrite accumulation, which have been associated with the production of carcinogenic substances.

1.3.4- Host for Genetic Engineering:

In recent years, the development of various expression systems for *S. carnosus* that allow for the production of heterologous proteins and the export of

these proteins to the cell surface has promoted the idea of employing this food-grade staphylococcus in the study of antigenic determinants and virulence factors of other pathogenic Gram-positive bacteria that are phylogenetically related, such as *S. aureus*. In particular, the possibility of creating live vaccines, i.e bacteria that express a single inactive antigenic marker in the cell surface, have attracted the scientific community to the advantage of S. *carnosus* as a host organism (Samuelson et al. 1995; Hansson et al. 2002).

Gotz et al. (Keller et al. 1983) were the first to study the production of plasmid vectors for cloning in the S. *carnosus* strain TM300, based upon the chloramphenicol resistance vector pC194 originally isolated from S. *aureus* (Figure 5), as well as the tetracycline resistance vector pMK148 (Horinouchi & Weisblum 1982). Subsequently, several chimeric versions of these vectors were produced specifically for the use in *S. carnosus* (Demleitner & Götz 1994). In particular, the vector pLipPS1, constructed with the lipase gene from *Staphylococcus hyicus*, and its corresponding membrane excretion pro-peptide signal, was used as the parental vector systems for the expression of membrane-bound and extracellular proteins in S. *carnosus*.

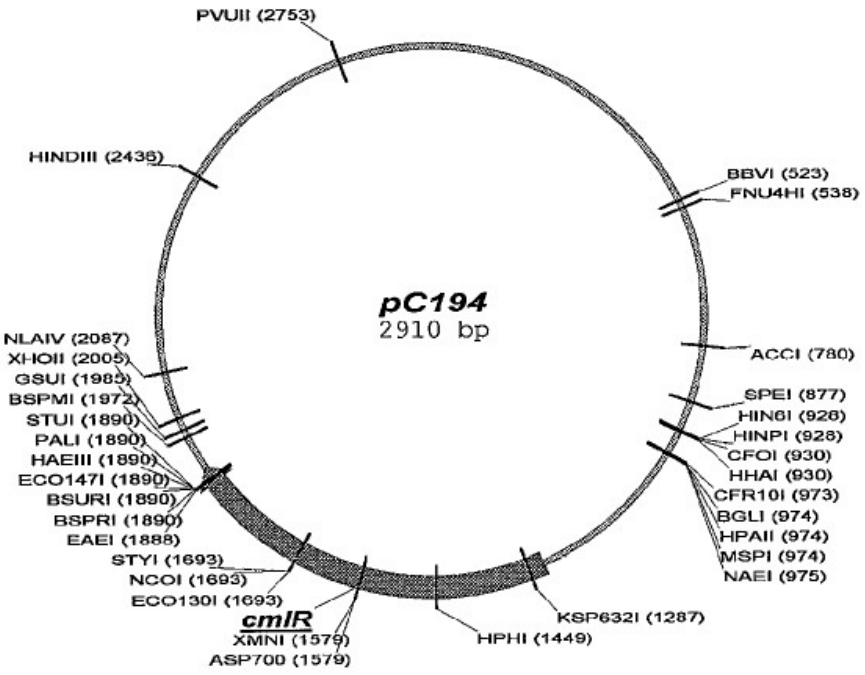


Figure 5- Map for the pC194 (ATCC® 37034™) plasmid. The area highlighted corresponds to the chloramphenicol resistance gene. The pCT20 vector was constructed by ligating a HindIII segment of pMK 148 into the unique HindIII site of pC194 (reproduced with the permission of the ATCC® website- www.**atcc**.org/**).**

Inspired by the development of new cloning and expression strategies directed at staphylococcus species, several research groups started to study the use of *S. carnosus* and S. *xylosus* as live vaccines. Samuelson et. al (Samuelson et al. 1995) created the first surface display expression vector by ligating a derivative of pLipPS1, pLipPS17, with the E. *coli* plasmid pRIT28. The resulting hybridized vector, which contained the lipase promoter and both the signal peptide and peptide sequences of the lipase gene, was linked with an ABP (albumin binding protein) sequence from the streptococcal protein G, and a cell wall anchoring sequences (from staphylococcal protein A). The resulting plasmid, pSPPmABPXM, was able to express, secrete and display ABP in *S. carnosus*, and induced a positive antibody response when tested through immunogold electron microscopy analysis. This vector was the first proof of concept for use of staphylococcal strains as vaccine display systems.

The positive immunization response achieved in this initial study led to the generation of "second generation" *S. carnosus* vaccine delivery systems that could induce an even greater immune response by co-displaying a fusion product from surface bound antigens and adhesion proteins of the targeted cells from the infected hosts. For instance, the co-expression of cholera toxin B subunit, which targets the molecule monosialoganglioside GM1 of mucosal surfaces, with an antigen on the surface of staphylococcal strains was shown to promote a stronger IgG response than antigen-presenting staphycocci that did not co-express adhesion proteins (Liljeqvist et al. 1997).

More recently, Williams et al. managed to create a derivative of pSPPmABPXM to allow for the intracellular expression of a broader range of heterologous proteins, by cleaving off the sequence involved in the export and anchoring of recombinant proteins to the cell wall (Williams et al. 2002).This shuttle vector, pNW21 (Figure 6), was then used to express the S. *aureus hysA* gene encoding for the putative virulent determinant hyaluronate lyase, which was shown to accumulate intracellularly in the cell.

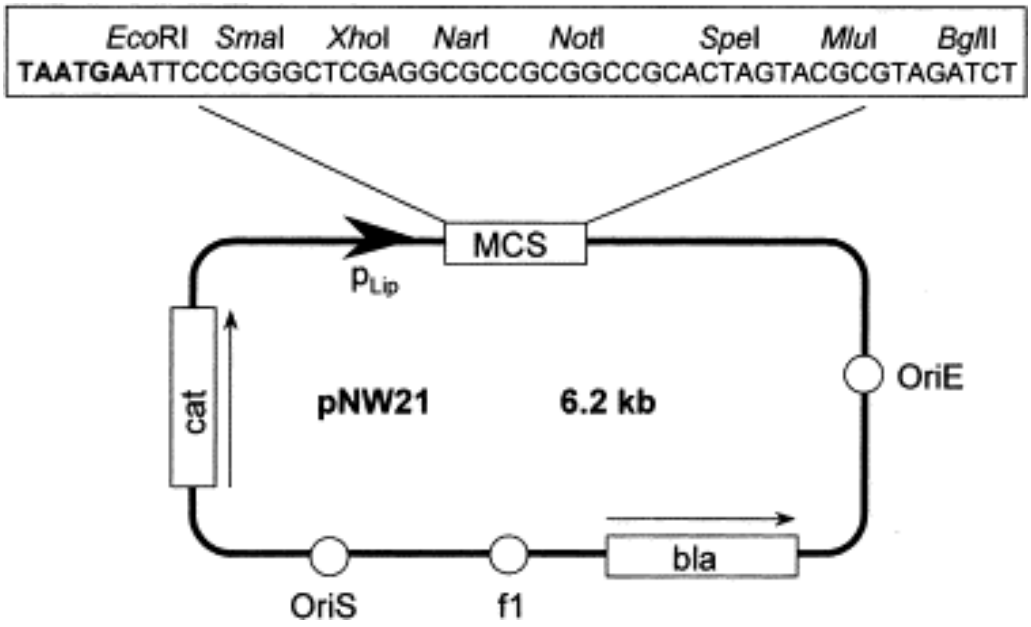


Figure 6- pNW21 vector, containing both origins of replication for *E.coli* and *Staphylococcus*, a multi-cloning site (MCS), a chloramphenicol resistance gene (cat), and a ampicillin resistance gene (bla) (Reproduced from *Plasmid*, vol. 47 (3), Williams et al., Expression of the S. aureus hysA gene in S. carnosus from a modified E. coli–staphylococcal shuttle vector, pg. 241-245, Copyright © 2002, with permission from Elsevier.)

To summarize, *S. carnosus* is a unique bacterial strain that shows promise as an efficient biocatalytic host. On one hand, its ability to grow in very acidic, water deprived environments with high salt concentrations, and the physical resilience of its cell wall, argue for an higher adaptability to adverse conditions when compared to more conventional Gram-negative strains such as *E. coli*. One can envision the possibility of applying *S. carnosus* biocatalytic systems to one-pot catalysis processes, in which the whole-cell system catalyzes one or more reaction steps in parallel to the chemical catalysis of the remaining steps in the process. In addition, *S. carnosus* has already been validated as safe-to-use, industrially relevant bacteria strain, which argues for a greater ease in scaling up and optimizing whole-cell biocatalytic processes performed with this strain. Finally, the development of genetic expression systems for staphylococcal species has increased the value of *S. carnosus* as a strain for biocatalytic expression by creating a precedent for the efficiently cloning and expression of different heterologous proteins. Therefore, we think there was an argument to be made for the use of *S. carnosus* as a host for biocatalysis, and the aim of this project was to pursue validation for this argument.

1.5- The Model Biocatalyst: Cyclohexanone Monooxygenase

When evaluating the efficiency of bacterial strains as biocatalytic hosts, the choice of biocatalyst is mainly dependent on the characteristics of the host. The general question is whether *S. carnosus* is a good whole-cell biocatalytic host, but this question is not very useful to define the important parameters that characterize *S. carnosus*. Rather, more specific questions need to be answered: can *S. carnosus* express a biocatalyst, does *S. carnosus* allow for biocatalysis to take place efficiently, and does *S. carnosus* offer a clear advantage over other

conventional species? These questions define what parameters need to be evaluated, and in consequence the best biocatalyst to use for testing these parameters.

Cyclohexanone monooxygenase (CHMO) from *Acinetobacter* sp. strain NCIB 9871 was chosen as the most suitable to answer these questions for a number of reasons. CHMO is a Baeyer-Villiger monooxygenase that requires the co-factors FAD and NADPH for catalysis, and therefore there is a clear motivation to use this enzyme in the context of whole-cell biocatalysis. In addition, some of the substrates and products catalyzed by this enzyme are of industrial relevance, but have been shown to be toxic to other bacterial strains in high concentrations. This poses a challenge to many whole-cell biocatalytic processes using CHMO, as efficient turnovers are largely dependent on the tolerance of the host strain.

More Importantly, CHMO-mediated biocatalysis has already been tested in a range of bacterial host strains, including *E. coli* and bakers yeast. On one hand, these precedents show that the enzyme can be efficiently expressed in different genetic environments. The body of literature on the use of CHMO in whole-cell systems is crucial to the characterization of the advantages and disadvantages of *S. carnosus* as a host when compared to other bacterial strains.

1.5.1- Baeyer-Villiger Monooxygenases and their Importance for the Biotechnology Industry

Baeyer-Villiger monooxygenases (BVMOs) are a sub-class of oxygenases that can perform Baeyer-Villiger oxygenation (BV oxygenation) reactions. The BV reaction was first discovered in 1899 by Adolf Baeyer and Victor Villiger (van Beilen et al. 2003) and describes the oxidation of linear, aryl and alkyl ketones into esters, and cyclic ketones into lactones.

In organic chemistry, BV oxygenation reactions are achieved by using nucleophilic compounds, the most common of which are peroxy-acids, which are involved in the nucleophilic attack on the carbonyl group of the substrate. The “Criegee intermediate” formed by this attack is labile and therefore will decay,

promoting the insertion of one oxygen atom between the carbonyl group and the adjacent carbon atom (Figure 7).

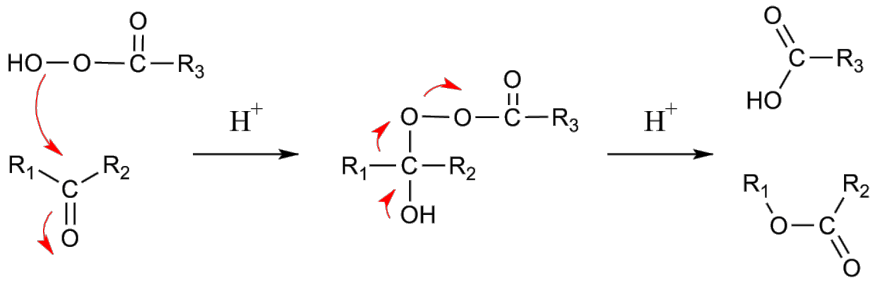


Figure 7- Baeyer-Villiger oxidation of a ketone, using a peroxy-acid as the nucleophile compound that attack the carboxyl group, forming a Criegee intermediate. This intermediate subsequently decays, and the nucleophilic oxygen is introduced into the carbon structure of the ketone (taken from http://en.wikipedia.org/wiki/File:Baeyer-Villiger_oxidation_(mechanism).png).

The industrial relevance of BV oxygenations relies on the fact that they are applicable to a large group of carbonyl compounds, and the reactions are highly stereo-specific and have a very predictable regiochemistry (Willetts 1997). An example of the importance of BV oxygenations in industry is the production of β-amino acids, which are important building blocks for peptides with enhanced resistance to proteolytic activity. These peptides, called β-peptides, can be formulated to work as drugs that are not so easily degraded or rejected by the body. In 2010 Rehdorf et al. reported the conversion of racemic N-protected β-amino ketones by different BV monooxygenases into “normal” β-amino acids, the normal products from the chemical catalysis of these ketones, as well as “abnormal” β-amino acids that cannot be obtained through chemical processes (2010). Another example is ε-caprolactone, the lactone product from the catalysis of cyclohexanone by CHMO, which is a constituent for highly specialized polymers such as polycaprolactone, which is used as a degradable biomaterial (Woodruff & Hutmacher 2010).

One of main challenges for the chemical industry in the application of BV oxygenation reactions is the use of oxidizing agents that are difficult to handle and

pose a health and environmental risk. In particular, peroxy-acids are shock-sensitive and explosive, making large-scale production a high-risk process. To counter-act this difficulty, alternative oxidative agents such as metal catalysts and organocatalytic compounds that use hydrogen peroxide and oxygen as oxidants have been established with moderate success (Kamerbeek et al. 2003). More importantly, peroxy-acids are powerful oxidative agents and can produce undesirable bi-products. The lack of enantioselectivity in chemical oxygenations remains the biggest issue for industry, and greatly contributes to the increasing interest in using biocatalytic processes as a more efficient, enantioselective and "greener" alternative.

The presence of enzymes capable of performing BV reactions was first reported in the 1940s in fungi that were involved in the biotransformation of steroids (Fraaije & Janssen 2007). The group of organisms capable of BV oxygenation was subsequently extended to Gram-positive and Gram-negative bacteria, plants, and shellfish (Kamerbeek et al. 2003). In these organisms, Baeyer-Villiger monooxygenases (BVMOs) were found to be involved in a broad range of functions, including toxin production, iridoids and steroids synthesis, and growth on aliphatic ketones and aromatic compounds.

The isolation of some of these proteins has led to the division of BVMOs into two discrete groups. Type I BVMOs were defined as flavin-dependent proteins composed of a single peptide chain that require NADPH and oxygen to perform oxygenation. These proteins are able to perform both the reduction of the flavin co-factor by NADPH and the BV oxygenation in the same active site (de Gonzalo et al. 2010). Examples of type I BVMOs include cyclohexanone monooxygenase (CHMO), which is the primary focus of our project, and phenylacetone monooxygenase (PAMO), the crystal structure of which has now been published (Fraaije & Janssen 2007). By contrast, type II BVMOs are composed of two distinct sub-units: a reductase subunit that uses NADPH to reduce the flavin co-factor; and a second subunit that performs the actual BV oxygenation reaction. The increased structural and functional complexity of this group might give insights into the rareness of the latter amongst the organisms so far studied (Fraaije & Janssen 2007).

1.5.2- The Discovery of Cyclohexanone Monooxygenase

Cyclohexanone Monooxygenase (CHMO) was first isolated by Chen *et al.* from *Acinetobacter* sp. strain NCIB 9871 (1988), and subsequently sequenced as an 542- amino-acid protein with the capacity of performing ketone-to-lactone conversions in an oxygen- and NADPH-dependent fashion. Iwaki et al. later corrected the sequence and the protein was identified as a 60.9 KDa monomer upon purification (Iwaki et al. 2002). The natural role of CHMO in *Acinetobacter* sp. NCIB 9871 is to degrade cyclohexanol into adipate (Mihovilovic et al. 2002). CHMO catalyzes the oxygenation of cyclic ketones through a mechanism very similar to that involved in chemical BV-reactions (Mihovilovic et al. 2003). In the first step, the bound flavin (FAD) is reduced by NADPH. The resulting enzyme-$NADP^+$ complex then reacts with oxygen and forms the flavin-peroxide nucleophile that performs the nucleophilic attack on the ketone substrate. In the final stage of the reaction, water is eliminated to convert the FAD into the original oxidized form, and $NADP^+$ is released thereby ending the reaction cycle (Figure 8).

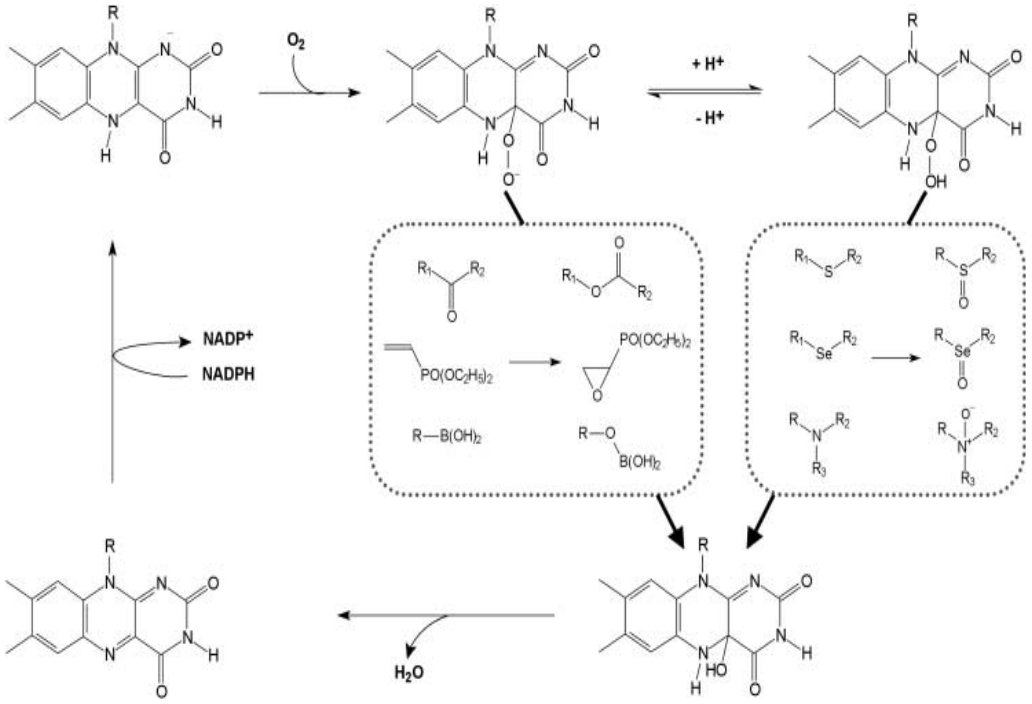


Figure 8- Mechanism of Baeyer-Villiger monooxygenase-mediated oxygenation showing the two distinct pathways which depend n the action of the reduced FAD as a nucleophile or electrophile (Reproduced from *Advanced Synthesis & Catalysis*, vol. 345 (6-7), Kamerbeek et al., Baeyer-Villiger Monooxygenases, an Emerging Family of Flavin-Dependent Biocatalysis, Copyright © 2003, with permission from John Wiley and Sons.).

The natural role of CHMO in *Acinetobacter* sp. NCIB 9871 is to degrade cyclohexanol into adipate, but it has since been tested against more than 100 non-natural prochiral and racemic substrates, achieving high regio- and enantioselectivity in most cases (Mihovilovic et al. 2002). These include mesomeric prochiral cyclohexanones with several side-chain substitutions, cyclobutanones, and prochiral bicyclic substrates with varying functional groups.

One of the most studied reactions is the conversion of bicyclo (3.2.0) hept-2-ene-6-one , because it demonstrates the capacity of CHMO to produce regio-enantiomers (Figure 9). In addition to producing two regioisomers in a 1:1 proportion from this ketone, CHMO also achieves a high level of enantiomeric purity for both lactones (Alphand et al. 2003).

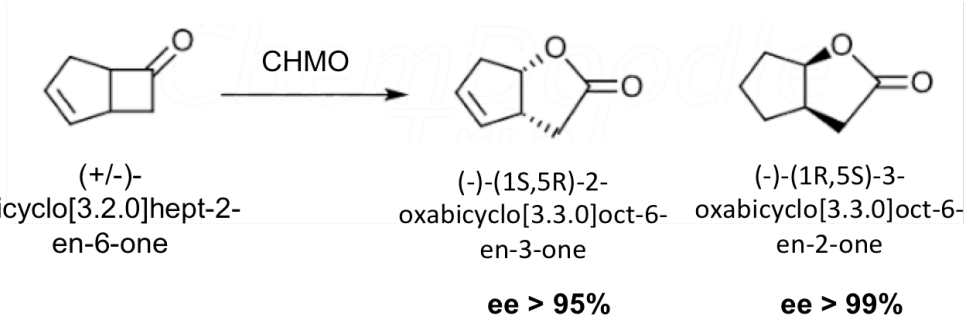


Figure 9- CHMO-mediated asymmetric oxygenation of racemic bicyclo (3.2.0) hept-2-ene-6-one 1 into its regioisomer lactones.

Due to its versatility, CHMO has attracted a great deal of research as well as commercial interest, which led to it being the most highly characterized type-I Baeyer-Villiger monooxygenase. It has also been one of the first in this class of enzymes to be used for biocatalysis i*n vitro* and *in vivo*, and has a proven potential as a potent biocatalyst in both cases.

1.5.3- Relation between Structure and Mechanism of CHMO

More recently, Ahmad Mirza et al. were able to produce two different conformational crystal structures of CHMO from *Rhodococcus* sp. strain HI-31 containing both co-factors FAD and $NADP^+$ (Figure 10) (Mirza et al. 2009). From the crystallography results, the protein was shown to be a single peptide organized into three domains: the flavin binding domain, which is contained within the termini of the protein sequence; a NADPH binding domain that in enclosed in the middle of the protein; and an helical domain composed of the segments within the NADPH domain which is believed to contain crucial residues that make up the substrate binding pocket. This modular arrangement showed high level of homology with the crystal structure of phenylacetone monooxygenase (PAMO), which had been previously solved (Malito et al. 2004).

The two different conformations of the CHMO are believed to represent different time points in the BV oxygenation reaction cycle, and comparison of the two has given insight into the role of the domains in the reaction mechanism. The active site has been pinpointed to a cavity near the flavin ring made up of the segments from the three domains. The trigger for the oxidation to occur was subsequently attributed to a slight conformational change in the NADPH domain that “pushes” the co-factor into a pocket located in close proximity to the FAD.

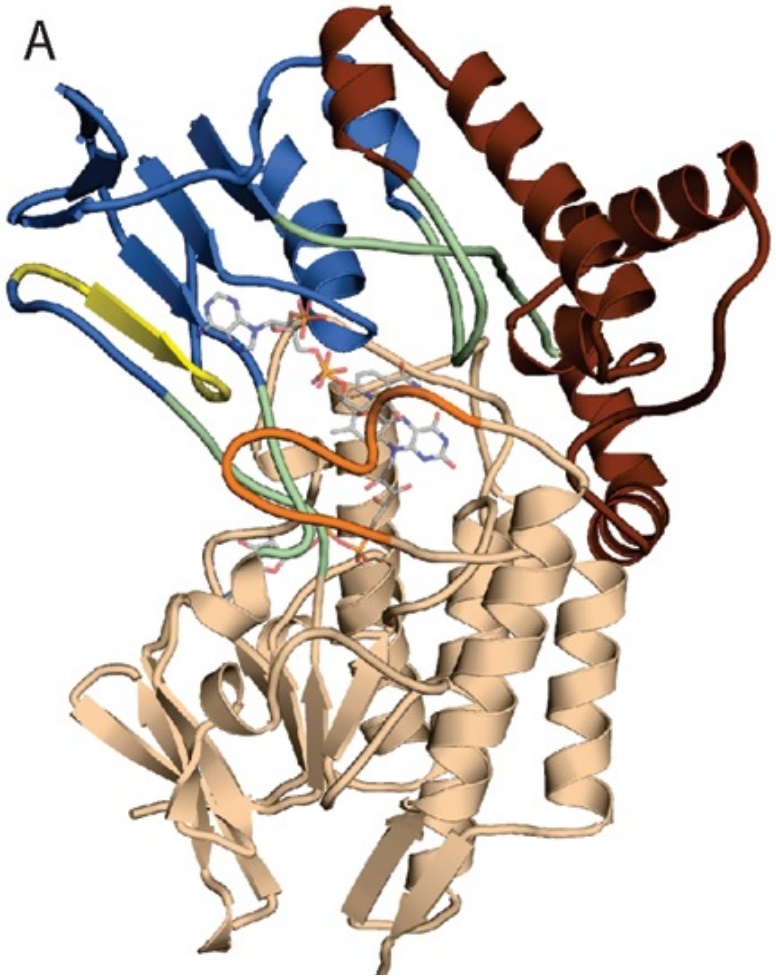


Figure 10- Structure of CHMO in the FAD and $NADP^+$ binding conformation. The dark-blue area represents the NADPbinding-domain; the wheat coloured area represents the FAD-binding domain. The figure shows the binding site enclosed in between the two domains, with the yellow and orange areas corresponding to the fexible loop regions also interacting with the compounds in the active site (reproduced with permission of Mirza et al., *Journal of the American Chemical Society*, Crystral Structures of Cyclohexanone Monooxygenase Reveal Complex Domain Movements and a Sliding Cofactor, Copyright © 2009, American Chemical Society.).

Some key residues involved in the reaction have also been identified. The arginine residue in position 239, which in contained in a flexible loop, interacts with NADPH directly and is probably involved in the insertion of nicotinamide ring into the pocket. Another residue that plays a crucial role in the lateral movement of NADPH is the tryptophan residue in position 492, also situated in a flexible region. This residue appears to play a role in the active site pocket, as deletion of the latter

results in the impairment of the enzyme activity (Mirza et al. 2009). Finally, an highly conserved region amongst type I BVMOs composed of the motif FXGXXHXXXWP at the start of NADPH domain is believed to coordinate the NADPH molecule into optimal position for the reduction of FAD during reaction cycle.

1.5.4- The Use of CHMO in Whole-Cell Systems

While the use of isolated-enzyme systems results in higher product yield and easier downstream processing, working with purified BV monooxygenases is not a straightforward process. Since these enzymes are NADPH- dependent, bioconversions at large-scale would require an efficient co-factor recycling system that would be both energy- and cost-intensive. In this case, the use of whole cells as biocatalytic hosts ends up being a more practical and economical option.

The wild-type *Acinetobacter* strain in which CHMO is naturally expressed is a class 2 pathogen, and thus using this organism for industrial applications is undesirable. To solve this problem, alternative, commercially viable strains of other species have been genetically engineered to express stable and active enzyme. Currently there is a large body of work documenting the construction and characterization of several whole-cell biocatalytic systems that allow for CHMO-mediated biocatalysis.

Cheesman et al. (2001) were able to express CHMO in both *E. coli* and *Saccharomyces cerevisiae*, obtaining a final protein yield of 7.3 mg for 3 L of yeast cultures and 12.9 mg for each liter of *E. coli* cultures. More recent studies were able to work on the optimization of CHMO expression in the *E. coli* system, obtaining yields of 60 mg- 80 mg of CHMO per liter of cells (Alphand et al. 2003). In a more biocatalytic-oriented study, Stewart et al. (1996) were also able to express the *Acinetobacter* wild type CHMO in bakers yeast and use the cells to perform BV oxygenations. Cyclohexanone was oxidized by the actively growing yeast cells into caprolactone with a reaction yield of 79%, and with minimal side-ractions (2.5%). The side- reduction of ketones by the indigenous metabolism of

the yeast seemed to be suppressed by the over-expression of CHMO. In spite of the importance of these studies in establishing micro-organisms suitable for the expression and bioconversion with this biocatalyst, large-scale studies were necessary to validate the industrial applications of these whole-cell systems.

Doig et al. (2001) were the first group to report and characterize the large-scale production of CHMO in *E. coli* (O'Sullivan et al. 2001). This was cloned using the vector pQR239, which expresses CHMO under an L-arabinose inducible system, and grown in volumes up to 300 L. Fermentation done on the 1.5 L scale, with optimal growth conditions, gave a final dry cell concentration of 5.5 g/L and the CHMO specific activity in the lysates of the cell broth was measured to be 630 U / gram of dried cells (1 U is the amount of protein that catalyzes the substrate-induced oxidation of 1 μmol of NADPH per minute). This represented a 50-fold increase of activity compared to the enzyme expressed in the native *Acinetobacter* strain. In the 300 L scale, similar cell concentrations were obtained when using the same growth conditions, but the specific CHMO activity measured in the lysates was significantly lower, at 500 U / gram of dried cells. This discrepancy was attributed to plasmid instability at the larger scale, where there is a greater percentage of cells that lose the protein expression vector in the course of the fermentation.

In a subsequent study, Doig et al. (2002) tested the biocatalytic potential of the *E. coli* strain expressing CHMO in the BV oxygenation of bicyclo(3,2,0)hept-2-en-6-one into its resulting lactone regiomers (Doig et al. 2002). Again, the optimal conditions were determined at the 1.5 L scale and subsequently scaled up to large fed-batch fermentations. On this scale, final cell density reached 5 g of dried cell weight per liter of broth, with a maximum whole-cell biocatalytic activity of 55 U/ gram of dried cells for conversion of the ketone substrate, while the enzyme specific activity in the lysates of the cell broth were measured to be 10 times higher (Doig et al. 2003). More recently, Baboo et al. (2012) used automated microwell technology to characterize the oxidation of 7 different ketones by the same *E. coli* strain, including cyclohexanone and Bicyclo[3.2.0]hept-2-en-6-one.

1.6- Limitations of CHMO Whole-cell Biocatalytic Systems:

The studies on CHMO expression and biocatalysis in bacterial and yeast strains were very important not only to characterize the successful implementation of CHMO as a biocatalyst in the context of whole-cell systems, but also to highlight the limitations of these systems. In the context of this PhD project, an understanding of the nature of these limitations is crucial to establish the parameters to study when determining the efficiency of alternative hosts. Therefore, it is worthwhile to briefly discuss both the cellular and enzymatic factors that hinder the performance of the CHMO whole-cell biocatalysis in *E. coli*.

(i) Optimal enzymatic pH and cell wall diffusion rates

One limitation that was briefly touched upon above was the lower activity observed for whole-cell biocatalysis of ketones when compared with the specific enzyme activity in lysates. This shortcoming in whole-cell biocatalysis was largely attributed to the difference between the pH of the intracellular microenvironment and the optimal pH for CHMO oxidation. Again, Doig *et al.* (2003) were the first to document this difference by showing that the activity rates of sonicated samples from bacterial cultures expressing CHMO sharply increased as the pH of the reaction buffer was shifted from a neutral to a more alkaline pH, with optimal activity being recorded for a pH of 9. The diffusion rate of the substrates across the membrane was also considered to contribute to the lower biocatalytic activities, albeit to a lesser degree.

Since the intracellular pH and membrane permeability are factors inherent and essential to the physiology of the *E. coli* bacterial host, these factors cannot be easily changed to suit the requirements of the biocatalyst without compromising the general metabolic activity of the host.

(ii) Oxygen availability

The availability of molecular oxygen to feed the BV oxygenation reaction can also be a limiting factor in the context of whole cell systems. This is due to the fact that, in addition to the stoichiometric requirement of oxygen for the CHMO-mediated oxidation reaction, cells also require oxygen for metabolic activity. Since the electron-transport chain, which is responsible for the metabolic energy production, exerts a higher catalytic pressure for the use of available oxygen in comparison to CHMO, BV oxygenation rates decrease in actively growing cells as the oxygen is primarily recruited to the metabolic pathways (Figure 11).

In a attempt to overcome this problem, Adam Walton & Stewart (2002; 2004) performed whole-cell CHMO-mediated oxygenations of cyclohexanone using non-growing cells. The reasoning behind the study was that non-dividing cells do not require large quantities of oxygen for metabolic purposes, and consequently will not exert such a high pressure on oxygen availability for BV oxygenation. Contrary to this assumption, the study reported that actively growing cells were more efficient biocatalytic units than the non-growing counterparts. However, the lower activity of non-growing cultures was offset by a higher cellular density, which resulted in an over-all 20-fold improvement of the volumetric productivity when operating in non-growing conditions.

The difference between biocatalytic efficiencies of growing and non-growing cultures could be tied to the differences in cell density. On the upside, resting cultures do not have a high oxygen consumption, but this is counteracted by the higher densities, which result in less oxygen availability per individual cell. Therefore, it seems the best solution to counter-act oxygen limitation is to prevent the creation of an oxygen limiting environment by close monitoring of fermenter DOT and oxygen tension during fermentation and bioconversion, as well as tightly controlling cell densities (Baldwin & Woodley 2006).

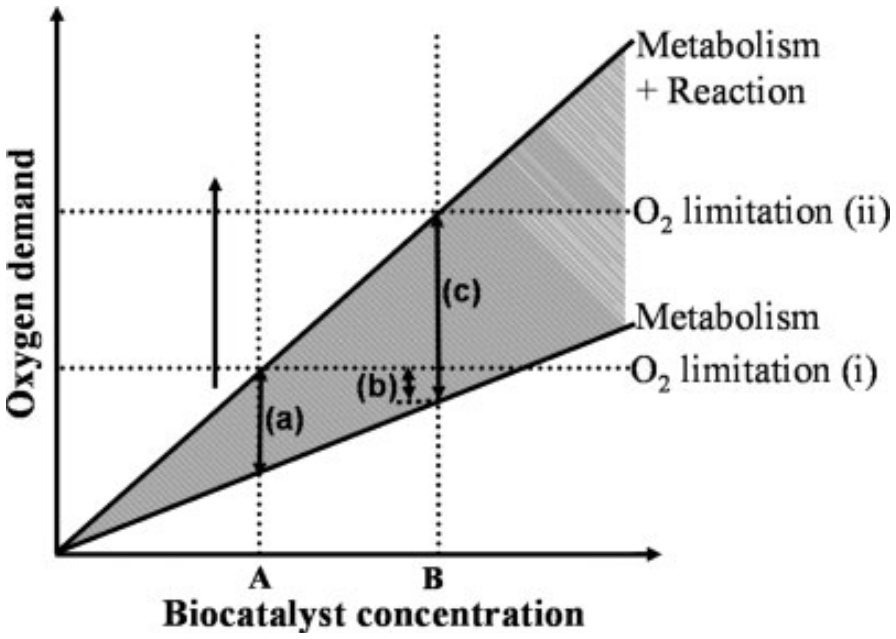


Figure 11- Model showing the affect of biocatalyst on oxygen demand. As biocatalyst concentration reaches point B, the oxygen demand is increased but the rate of biocatalysis wont increase from (a) to (b) unless oxygen supply is increased to counteract the overall oxygen demand (ii). This increase will result in an increase of biocatalysis rate to (c) (Reproduced from *Biotechnology & Bioengineering*, vol. 95(3), Baldwin et al., On oxygen limitation in a whole cell biocatalytic Baeyer-Villiger oxidation process, Copyright © 2006, with permission from John Wiley and Sons.).

(iii) Substrate/Product toxicity

Biocatalyst inhibition through substrate and product toxicity is the most relevant limitation when scaling up a whole-cell biocatalytic process, as large quantities of the reactants will kill the host cell. In the case of the CHMO-expressing *E. coli* system, several substrates and their respective products have been shown to severely hinder the biocatalytic process when concentrations rise above very low concentration thresholds. In particular, the substrate bicyclo[3.2.0]hept-2-en-6-one greatly inhibits the biocatalytic process at concentrations above 0.4 g/L, while its lactone products, (−)-(1*S*,5*R*)-2-Oxabicyclo[3.3.0]oct-6-en-3-one and (−)-(1*R*,5S)-3-oxabicyclo[3.3.0]oct-6-en-2-one, kill off the activity of the cells above concentrations of 5 g/L (Doig et al. 2003).

It is now generally accepted that the toxicity exhibited by some CHMO substrates and products is due the lipophilic character of these compounds, which will accumulate in the lipid membrane of the cell and cause severe changes in the structure and function (Sikkema et al. 1995). In extreme cases, this accumulation will lead to loss of membrane integrity, thereby compromising the viability of the cell.

The aggressive nature of many of the CHMO substrates and products characterized thus far towards the bacterial host pose a challenge to the application of CHMO whole-cell biocatalytic processes in an industrial setting, where high yields of product are expected. The toxicity of these compounds demands techniques that allow for a tight control over the accumulation of substrate and product in the reaction medium. One of the proposed options to tackle this issue is the application of *in situ* product removal (ISPR) systems, such as resins, during the biocatalytic process (Figure 12) (Lye 1999).

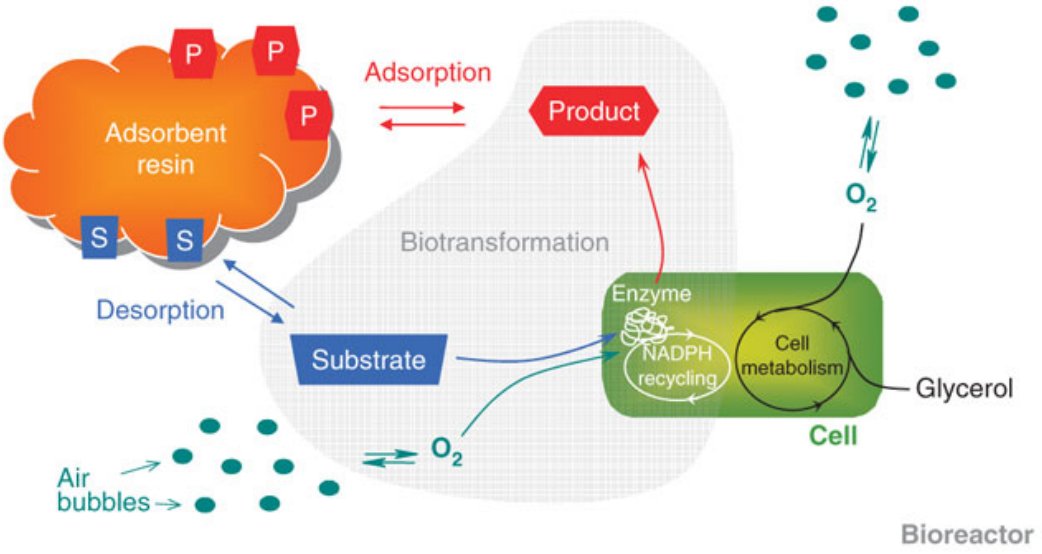


Figure 12- Schematics of whole-cell bioconversion using an absorbent resin for product feeding and substrate removal (Reproduced by permission from Macmillan Publishers Ltd: *Nature Protocols,* Hilker et al. 2008, Copyright © 2008.).

ISPR systems involve resins that interact with the reactants through covalent or non-covalent interactions and that can discriminate between the

stereoisomers and molecules with different functional groups. The use of resins that can bind specifically to the substrates and products is an appealing idea, as they can work as simultaneous substrate feeding and product recovery system, with the ready-made product displacing the bound substrate. In this manner, the reactants concentrations can be maintained below the toxicity threshold in the reaction medium. An added benefit of this system is the simplification of the subsequent product recovery process, which can be done by extracting the resins from the medium. Finally ISPR systems can also be used to shift unfavorable reaction equilibriums by maintaining concentration gradient directed toward product formation. In turn, the artificial gradient created by the ISPR system protects the unwanted degradation of substrates and products by decreasing the residence time of these compounds in the intracellular space of the host and the reaction medium (Woodley et al. 2008). So far, only a couple of studies have focused on the implementation of IPSR systems in CHMO-mediated biocatalytic processes. In 2001 Simpson et al. were able to successfully apply molecularly imprinted polymeric (MIP) resins for *in situ* feeding of bicyclo(3,2,0)hept-2-en-6-one and recovery of the respective regioisomers. In this study, two commercially available resins, XAD-4 and Optipore L-493, were added to a preparative scale vessel with cell concentration of 17 g/L, and allowed for the bioconversion of the ketone substrate at concentrations up to 20 g/L. This represents a considerable improvement compared to the 0.4 g/L ketone limit established for bioconversions without ISPR systems. By adjusting the ratio of resin to substrate, this process was subsequently scaled up to allow for the conversion of 1 kilogram of bicyclo(3,2,0)hept-2-en-6-one at a concentration, with a final lactone yield of 60% (Hilker et al. 2005). Despite the significant loss of product due to the non-optimal extraction process, the ketone and lactone concentrations were maintained below the toxicity threshold, thus ensuring that the cell viability was not compromised during the biocatalytic process.

1.7-Conclusions

1.7.1- The purpose of the project

Throughout this chapter I have attempted to give a broad picture of the current state of art in biocatalysis in order to provide a backdrop to this project.

Indeed, it should be now clear that whole-cell biocatalysis is a concept that still stirs the imagination with its massive potential. The advent of Synthetic Biology and the development of techniques that allow for a straight-forward, rational and systematic approach to cell engineering have promoted a renewed interest in the power of prokaryotic and eukaryotic organisms as biological factories.

However, genetic and metabolic engineering is only part of the solution. The creation of engineered iterations of known and industrially proven cellular strains is a powerful tool to increase flexibility and improve yield, but the changes that are being artificially introduced are not trivial and might create a burden for the host system, while not necessarily covering for limitations that are host-specific.

It is narrow-minded to think that the production of efficient whole-cell systems can only be achieved through genetic engineering. This argument would be neglecting one of the reasons why whole-cell biocatalysis can be advantageous, and that is the vast diversity that nature offers. There is a very big repertoire of bacterial and yeast species that have not yet been studied in this context. It is therefore premature to assume that the current cell-systems being used in the biocatalytic industry are the best suited for whole-cell biocatalysis.

An effort should be made by researchers (that are not directly limited by profit margins and therefore can take risks) to step out of the comfort zone and look for alternative strains that offer clear advantages for biocatalytic processing. The prospecting of natural alternatives to the conventions of whole-cell biocatalysis could, in the long run, be more a cost- and time- effective way to achieve efficient cell-biocatalysts, with the valuable side-effect of increasing the repertoire of biocatalysts and strains to choose from.

This conviction is at the center of this project, but it is not rejecting the value of synthetic biology and genetic engineering as essential pieces in this jigsaw that is designing an efficient biocatalytic process. Indeed, some level of genetic engineering must always be done in order to introduce a new biocatalytic system into a host, or to enhance the native capability of the host to perform biocatalysis.

1.7.2- Challenges of the Project

Searching for and accessing alternative hosts for biocatalysis is not a risk-free procedure. The huge library of known microorganisms that have not been used in the context of biocatalysis makes for a difficult choice when deciding the group of alternative hosts to study.

In addition, the myriad of factors involved in biocatalytic processes make up for a complex system of interactions that hampers the ability to predict strain performances, more so if the strain has not been characterized previously.

In the case of this project, *S. carnosus* was chosen because it showed promising phenotypical and biochemical features, such as an high tolerance to harsh environments, and the potential to accept compounds more readily because of the different membrane composition. However, there is no guarantee that these native features actually have a positive effect on whole-cell biocatalysis.

The novelty in using *S. carnosus* as biocatalytic host creates its own challenge, as there are no known examples or procedures to follow in the literature. As a result, the choice of conditions in which S. carnosus can work as an efficient host will rely heavily in educated guesses. Another challenge is that there are very few genetic tools available to introduce novel biocatalytic processes into *S.carnosus*. Most of vectors designed for use in *S. carnosus* were designed for diagnostic purposes, and as a result are based on the expression of heterologous proteins that are then secreted and displayed on the surface of cells. These vectors are not useful when the biocatalytic process occurs in intracellular space. In addition, only a couple of promoters have been studied in these genetic

systems, and therefore we have very limited knowledge of how to modulate and enhance biocatalyst expression in *S.carnosus*.

Chapter 2: Material and Methods

2.1- Materials

2.1.1- Media, Buffers, and Selective Antibiotics

Most buffers used for the plasmid purification and gel extraction were part of the kits from QIAGEN:

- P1 - resuspension buffer: 50 mM Tris-Cl, pH 8.0; 10 mM EDTA; 100 µg/ml RNase A.
- P2- lysis buffer: 200 mM NaOH, 1% SDS (w/v)
- N3- neutralization buffer: confidential
- P3-neutralization buffer: 3.0 M potassium acetate, pH 5.5
- EB - solubilisation buffer : 50 mM Tris-Cl, pH 8.0
- PE - wash buffer: confidential
- QC- wash buffer: 1.0 M NaCl; 50mM MOPS, pH 7.0; 15% isopropanol
- QBT- Equilibrium buffer: 750 mM MOPS, pH 7.0; 15% isopropanol; 0.15% Triton X-100
- QF- Elution buffer: 1.25 M NaCl; 50 mM Tris-Cl, pH 8.5; 15% isopropanol
- PB- binding buffer: confidential
- QG –solubilisation buffer: confidential

For protocols of cell lysis and DNA extraction from *S. carnosus* cultures, the following specific buffers and chemical solutions were used:

- Cell lysis buffer **L1**: 50 mM Tris-HCl buffer, pH7.4; 30 - 60 mg/ml Lysozyme from chicken egg white (Sigma-Aldrich)
- Cell lysis buffer **L2**: 50 mM Tris-HCl buffer, pH7.4; 60 mg/ml Lysozyme from chicken egg white (Sigma-Aldrich); 1 µg/ml Lysostaphin from *Staphylococcus staphylolyticus* (Sigma-Aldrich)
- Cell lysis buffer **L3**: 50 mM Tris-HCl buffer, pH7.4; 1 µg/ml Lysostaphin from *Staphylococcus staphylolyticus* (Sigma-Aldrich)
- Solubilisation buffer: 50 mM Tris-Cl, pH 8.0
- SDS solution: 3% (w/v) SDS; 0.2 M NaOH
- 3 M sodium acetate, pH 4.8
- 7. 5 M ammonium acetate with 0.5 mg/ml Ethidium bromide (Sigma-Aldrich)
- Phenol:Chloroform:Isoamyl Alcohol 25:24:1 Saturated with 10 mM Tris, pH 8.0, 1 mM EDTA (Sigma-Aldrich)
- Ethanol (99%)
- Ethanol (70%)
- Isopropanol (98%)

For DNA electrophoresis, the following buffers were used:

- TBE Running buffer: 89 mM Tris-Borate, 2 mM EDTA buffer 1X (Sigma-Aldrich)
- Stacking gel: TBE buffer 1X (Sigma-Aldrich); 0.7-1% agarose (Sigma-Aldrich) ; 5 to 10 µl Ethidium bromide from a stock solution of 1% (w/ v) (Sigma-Aldrich).

For SDS-PAGE, the following buffers were used:

- 50 mM phosphate buffer saline (PBS) (pH 7.4): containing 3.5X of 1 MPotassium Phosphate (dibasic) and 1X of 1 M Potassium Phosphate (monobasic).
- Protein loading buffer (2X) stock solution: 0.8 g of SDS; 0.4 ml of β-mercaptoethanol (Sigma-Aldrich); 8 mg of Coomassie Brilliant Blue G-250 (Thermo Scientific); 4 ml of 100% glycerol for every 10 ml of solution.

- Protogel 30% (National Diagnostics): acrylamide/ methylene bisacrylamide solution (37.5:1 ratio).
- 4X Resolving buffer (National Diagnostics): gel of 0.375 M Tris-HCL, and 0.1% SDS, pH 8.8.
- ProtoGel Stacking buffer (National Diagnostics): gel of 0.125 M Tris-HCL and 0.1% SDS, pH 6.8.
- *N*,*N*,*N*′,*N*′-Tetramethylethylenediamine (Sigma-Aldrich)
- 10 % (w/v) ammonium persulfate

Growth media were used for the inoculation and fermentation of cell cultures, production of selective agar plates, and for culture storage:

- Nutrient Broth no. 2 (Oxoid) agar: La-Lemco powder 10g/L; Peptone 10g/L; sodium chloride 5g/L, pH 7.5; 2% (w/v) Agar granulated (BD) in water.
- B2 Medium: 2.5 % (w/v) Yeast Extract (DB); 2.5% (w/v) NaCl (DB); 1% (w/v) Casein Hydrolysate (DB), 0.5% (w/v) Glucose (DB), 0.1% (w/v) K_2HPO_4 (DB) in water.
- LB medium: 10 g/L Tryptone (DB); 5 g/L Yeast Extract (DB); 10 g/L NaCl (DB)
- Nutrient Broth Agar: 25 g/L Nutrient Broth no. 2 (Oxoid), 2% (w/v) DIFCO™ Agar (DB) in water.
- -80 ^{o}C storage solution: 10-50 % Glycerol in water.

2.1.2- Restriction and Ligation Enzymes

A list of single and combinations of restriction enzymes used to create the restriction maps and digestion for ligation together with the buffers used in reaction can be found below. The reaction buffers used are in accordance with the guidelines in the NEB website (http://www.neb.uk.com/):

- *Eco*RI (NEB) ; *Eco*RI reaction buffer 1X (NEB), containing 100 mM Tris-HCl buffer (pH 7.5), 50 mM NaCl, 10 mM $MgCl_2$, and 0.025% Triton® X-100.
- *Xho*I (NEB); NEBuffer 2 1X (NEB), containing 10 mM Tris-HCl buffer (pH 7.9), 50 mM NaCl, 10 mM $MgCl_2$, and 1 mM DTT.
- *Pst*I (NEB); NEBuffer 3 1X (NEB), containing 50 mM Tris-HCl buffer (pH 7.9), 100 mM NaCl, 10 mM $MgCl_2$, and 1 mM DTT.
- *Sac*I (NEB); NEBuffer 1 1X (NEB) containing 10 mM Bis-Tris-Propane-HCl buffer (pH 7), 10 mM $MgCl_2$, and 1 mM DTT; BSA 1X (NEB), containing 20 mM KPO_4 buffer (pH 7), 50 mM NaCl, 0.1 mM EDTA, and 5% glycerol.
- *Nde*I (NEB); NEBuffer 4 1X (NEB), containing 20 mM Tris-Acetate buffer (pH 7.9), 50 mM Potassium Acetate, 10 mM Magnesium Acetate, and 1 mM DTT.
- *EcoR*I (NEB); *Xho*I (NED); *Eco*RI reaction buffer 1X (NEB)
- *Eco*RI (NEB); *Hind*III (NEB); *Eco*RI reaction buffer 1X (NEB)
- *Eco*RI (NEB); *Pst*I (NEB); *Eco*RI reaction buffer 1X (NEB)
- *Eco*RI (NEB); *Nde*I (NEB); *Eco*RI reaction buffer 1X (NEB)
- *Xho*I (NEB); *Hind*III (NEB): NEBuffer 2 1X (NEB)
- *Xho*I (NEB); *Nde*I (NEB); NEBuffer 2 1X (NEB)
- *Xho*I (NEB); *Pst*I (NEB); NEBuffer 3 1X (NEB)
- *Pst*I (NEB); *Hind*III (NEB); NEBuffer 3 1X (NEB)
- *Sac*I (NEB); *Nde*I (NEB); NEBuffer 4 1X (NEB)
- *Sac*I (NEB); *Hind*III (NEB); NEBuffer 2 1X (NEB)

The ligase used for the plasmid ligations were the T4 ligase from NEB with the corresponding reaction buffer 10X (1X containing 50 mM Tris-HCl buffer pH 7.5, 10 mM $MgCl_2$, 1mM ATP, and 10 mM DTT), and the Quick-Stick Ligase (Bioline), with the corresponding Quick-stick Buffer 10X (1X containing 66 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 1 mM Dithiothreitol, 1 mM ATP, and 7.5% Polethylene glycol (PEG 6000)).

2.1.3- Strains and Vector Constructs

E. coli DH5α (F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17(r_K^- m_K^+), λ–) and TOP10 (F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ-) were used for vector purification, production of competency and vector transformation. This strains were proived by John Ward's lab and commercially. *S. carnosus* TM300 (wild-type) was used throughout the transformation and protein expression studies. This strain was provided by the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures (DSMZ).

E. coli plasmids used for ligations and restriction digests were pNW21 (created in John Wards lab) and pTTQ18 (Stark 1987). The *S. carnosus* vector used for the ligation reactions was the pCT20 (a variant from pC194), ordered from DSMZ (German Resource Centre for Biological Material). *E. coli* vector pJ201 (supplied by John Ward's lab) was used as the source for the gentamicin resistance gene. The pQR239 vector (provided by Frank Baganz's lab) was used as the source for the CHMO gene.

2.2- Molecular Biology Techniques

2.2.1- Plasmid DNA Extraction

Extraction and purification of plasmid DNA from *E. coli* strains was achieved with spin columns miniprep kits (Qiagen). This kit uses lysis and ionic buffers to break the cells and columns that bind to the plasmid DNA through gravitational force, separating it from the cell debris. The following protocol was used:

- Colonies of *E. coli* clones were grown in Falcon tubes containing 5ml of NB2 or TB medium and the necessary selective antibiotic (100 µg/ml of ampicillin; or 15 µg/ml of gentamicin) for 16 hours (overnight), in a Inova 4330 (New Brunswick) shaking incubator set with a constant temperature of 37^0C and a shaking speed of 250 rpm.
- The resulting bacterial cultures were collected from the growth medium by centrifugation using a centrifuge 5810R (Eppendorf) (5 minutes (min), 16 639 g, 4 ^{0}C) with a rotor for 50 ml falcon tubes, and resuspended in 250 µl of P1 buffer, which were subsequently transferred to 1.5 ml sterile eppendorf tubes.
- Cell lysis was induced by addition of 250 µl of P2 buffer.
- 350 µl of N3 buffer was subsequently added after 5 minutes of incubation at room temperature to neutralize the lytic reaction, and to induce aggregation of large macromolecules that could inhibit subsequent DNA purification.
- The neutralized samples were centrifuged on a bench top accuSpin™ Micro centrifuge (1 min, 17 000 g) in order to fractionate the soluble content from the aggregated cell debris, and the resulting supernatant was loaded unto a solid phase extraction column that binds to DNA between 100bp and 20 Kb.
- The supernatant samples were washed through the columns using centrifugal force (1 min, 17 000 g), and the resulting flow-through was discarded. The plasmid DNA bound to the columns was subsequently washed twice, first with 500 µl of PB buffer, followed by 750 µl of PE buffer. These washing steps were done to clean the DNA samples from impurities, such as protein and lipid structures, that might have attached to the columns during the first flow-through of the supernatants. In both steps, centrifugal force (1 min, 17 000 g) was used to wash the buffers through the solid phase of the columns.
- After the washing steps, the plasmid DNA samples were eluted in 50 µl of EB buffer from the columns into 1.5 ml eppendorf tubes by centrifugal force (1 min, 17 000 g). Final DNA concentrations were measured using a Nanodrop 2000 spectrophotometer (Nanodrop) at a wavelength of 260 nm.

For *S. carnosus*, the kits mentioned above were not very efficient, recovering less than 5ng/ul of plasmid DNA per 5ml of overnight culture. Instead,

two different methods were used for plasmid extraction from the staphylococcal cultures.

The first protocol was based on a previously developed methodology by Sullivan et al. (O'Sullivan & Klaenhammer 1993), which involved a phenol-chloroform gradient to extract the plasmid from cells debris. The following modified extraction protocol was used:

- Colonies of *S. carnosus* strains containing plasmid constructs were grown in Falcon tubes containing 5 ml of B2 or NB2 medium with the necessary selective antibiotic (15 µg/ml of ampicillin; or 15 µg/ml of gentamicin) for 16 hours, using a shaking incubator set with the temperature of 37^0C and shaking speed of 250 rpm.
- Resulting cultures were collected from the growth broth via centrifugation (5 min, 16 639 g, 4 ^{0}C), and subsequently resuspended in 250 µl of a 25% sucrose solution containing 30 to 60 mg/ml of lysozyme. The resuspended samples were incubated for 15 minutes in a 37 ^{0}C water-bath.
- 250 µl of the Qiagen lysis buffer P2 was added to samples, which were subsequently incubated for a further 7 minutes at room temperature.
- The lysis reaction was stopped by adding 300 µl of 3 M sodium acetate (pH 4.8), which was cooled to a temperature of 4 ^{0}C prior to addition. In addition to neutralizing the lytic process, the sodium acetate solution also induces the aggregation of large cell fragments, in a similar fashion to the N3 buffer.
- After aggregation, the samples were centrifuged (10 min, 17 000 g) and the supernatants were transferred to 1.5ml sterile Eppendorf tubes, to which 650 µl of 98% isopropanol was added.
- After mixing, the samples were centrifuged for a further 15 minutes, resulting in a white pellet that was resuspended in a solution containing 320 µl of sterile water, 200 µl of 7.5 M ammonium acetate with 0.5 mg/ml of ethidium bromide, and 350 µl of phenol/chloroform.
- The samples were vigorously mixed to produce homogenous solutions, which were centrifuged (1 min, 17 000 g) in order to promote DNA separation from the other soluble debris through a solvent gradient.

- The upper phase of the gradient was collected and transferred to a new sterile 1.5 ml eppendorf tube, to which 1 ml of ethanol was added to wash the DNA.
- After another round of centrifugation (1 min, 17 000 g), pellets of DNA, which formed in the bottom of the tube, were washed a second time with 70% ethanol, subsequently dried in a 55 ^{0}C incubator for 15 min.
- DNA samples were finally resuspended in 50 µl of EB buffer. Concentrations were measured using a nanodrop spectrophotometer at 260 nm.

The second protocol used for plasmid extraction from staphylococcal cultures was a simplified version of the phenol-chloroform extraction protocol, in which the solid phase QIAGEN column technology was used for DNA purification after pre-treatment of cultures with a lysozyme/lysostaphin solution:

- Overnight cultures were collected via centrifugation (5810R, 5 min, 16 639 g, 4 ^{0}C), resuspended in 250 µl of one of three different lysis buffers (L1, L2, and L3), and transferred into 1.5 ml sterile eppendorf tubes. The resulting samples were incubation for 1 or 2 hours in a 37 ^{0}C water-bath.
- 250 µl of P2 lysis buffer was subsequently added to samples, which were incubated for a further 5 minutes at room temperature.
- Lysis neutralization and aggregation of unwanted cell debris was induced by adding 350 µl of N3 buffer.
- After centrifugation (1 min, 17 000 g) , the resulting supernatants were loaded unto Qiagen spin columns, and the plasmid DNA was subsequently purified the Qiagen purification protocol described for *E. coli*.

2.2.2- DNA assembly techniques

(i) DNA restriction and purification

The different vector constructs were produced by site-specific ligation of endonuclease-restricted fragments. For most ligation protocols done in this project, the DNA fragments were cut with different endonucleases (REs) at the ends, to allow for directional ligation of the different fragments.

The restriction protocols were conducted prior to the ligation reactions to either extract DNA fragments from plasmids, or to produce the appropriate ends in amplified products and vectors. Restriction digest were also performed after ligation to check for ligation products and produce diagnostic restriction maps.

In all cases, restriction reactions were performed by adding a volume of endonuclease below 5% of total reaction volume into a 1.5 ml sterile Eppendorf tube containing 10- 50 µl of DNA. In the case of reactions with a mix of different endonucleases, the combined volume for the enzymes did not exceed 5% of the total volume of reaction, above which concentration the solution in which the endonucleases are stored becomes inhibitory to the restriction reactions. In addition, 1/10 volume the corresponding reaction buffer (10X) (see section 2.1.2), was also added to provide the required co-factors for reaction to occur. Finally, sterile water was added to the reaction mix to make up for the rest of the total reaction volume, if necessary. As an example, a 20 µl restriction digest of 10 µl of plasmid DNA would be made up in the following fashion:

$$Reaction\ mix\ (20\mu l) = 10\mu l\ \text{plasmid DNA} + 1\mu l\ \text{RE} + 2\ \mu l\ \text{RE buffer}\ (10X) + 7\ \mu l\ H2O$$

Restriction reactions were incubated in a 37 ^{0}C water-bath for 2 to 4 hours, depending on the restriction rates of the different REs, and subsequently stopped by heat-shock treatment (10 min, 65 ^{0}C heat-block). All the endonucleases used in this study produced single-stranded overhangs upon excision of the target restriction sites.

Prior to ligation, restricted DNA samples have to be separated from the inactivated reaction mix, which otherwise could interfere with the ligation process. This was done either by gel extracting the DNA fragments from an agarose gel, or in the case no fragment fractionation was needed, by using Qiagen PCR purification kit. These kits are optimized to purify amplified DNA fragments from the enzymatic mixes used during polymerase chain reaction (PCR) protocols. Therefore, the kits are also suitable to wash away the mix of inactivated REs and reaction buffer impurities from the restricted DNA material. For purification by Qiagen purification kit, the following protocol was used:

- 5 volumes of PB buffer were added to 1 sample volume, contained within a 1.5 ml sterile Eppendorf tube. For example, for 50 µl of DNA restriction sample, 250µl of PB buffer were used.
- The resulting mix was loaded unto a Qiagen purification column, which similarly to the Qiagen miniprep columns contains a solid phase that binds to plasmid DNA in solution through centrifugal force.
- The mix was washed through the column by centrifugation (1 min, 13 000 rpm), with the resulting flow-through being discarded.
- The column was subsequently washed with 750 µl of PE buffer to wash away cell debris impurities that might have been absorbs into the solid matrix of the column.
- The plasmid DNA was eluted in 20 µl of EB buffer by centrifugal force (1 min, 17 000 g), and its final concentration was checked through a nanodrop spectrophotometer set for a wavelength of 260 nm.

The process of plasmid DNA purification through the use of the Qiagen purification kit not only generated samples of high purity, but also resulted in a reduction of the sample's volume, consequently increasing the final concentration

of the DNA product. In the case of the example given above, an initial sample volume of 50 µl was reduced to a final volume of 20 µl. By contrast, the alternative method of DNA purification, gel extraction, often results in a loss of final sample concentration, but is a requirement for the purification of heterogeneous samples. This protocol is described in the next section.

(ii) Ligation of purified fragments

Ligation protocols were performed with a variety of different commercially available DNA ligases and operating conditions, which were partly dependent on the type of ligase. T4 ligases provided from New England Biolabs and Invitrogen were used alternately in most ligations protocols involving fragments with single stranded overhangs. In addition, the Quick-Stick Ligase, from Bioline, was also used as a faster alternative to the standard protocols in ligation steps that did not require a long incubation stage.

Prior to ligation, the different restricted fragments were mixed together following specific molar ratios between the fragments. Depending of these ratios, the volumes of the different fragments to be added together were calculated using the following equation (1):

$$ng\ insert = \frac{ng\ of\ vector * size\ of\ insert}{size\ of\ vector} * molar\ ratio\ (vector{:}\ insert) \quad (1)$$

Where insert refers to the smaller of the fragments, and vector corresponds to the larger fragment. Alternately, vector often referred to the DNA fragment containing the selection and replication genes. For instance, in the case of the ligation between pTTQ18 and the synthetic linker (see Chapter 4, section 4.4), the former is considered to be the vector due to its larger size, and the fact that it contains the ampicillin resistance gene and the origin of replication. The sizes of

the different fragments are expressed as kilo-bases (Kb), and the amounts referred to in nano-grams (ng).

Various molar ratios were tested during the ligation protocols in order to find the optimal ratio between the different fragments. For ligations involving DNA fragments with single-stranded overhangs, the molar ratios 1:1 (vector/insert), 1:3 (vector/insert) and 1:6 (vector/insert) were used. For blunt-end ligations, higher ratios of 1:6 (vector/insert) and 1:10 (vector/insert) were used in order to increase the probability of ligation between the phosphorylated ends of different fragments, as opposed to re-ligation of homologous molecules.

The components of the different ligation reactions were as follows:

- For the T4 ligase from NEB:

$$Total\ volume\ (10\ \mu l) = x\ \mu l\ of\ vector + y\ \mu l\ of\ insert + 1\ \mu l\ Reaction\ buffer\ (10X) + 10 - (1 + x + y)\mu l\ of\ H2O$$

- For the T4 ligase from Invitrogen:

$$Total\ volume\ (20\ \mu l) = x\ \mu l\ of\ vector + y\ \mu l\ of\ insert + 5\ \mu l\ Reaction\ buffer\ (4X) + 20 - (5 + x + y)\mu l\ of\ H2O$$

- For the Quick-Stick ligase from Bioline:

$$Total\ volume\ (10\ \mu l) = x\ \mu l\ of\ vector + y\ \mu l\ of\ insert + 1\ \mu l\ Reaction\ buffer\ (10X) + 10 - (1 + x + y)\mu l\ of\ H2O$$

Volumes of the vector and insert fractions were derived from the molar ratios of reactions and concentrations of the different DNA samples. Ligation reactions were routinely incubated in a 37 ^{0}C water-bath for 1 to 2 hours, with the

exception of reactions involving the Quick-Stick ligase, which were incubated for 5 minutes. Alternatively, T4 ligase reactions were also incubated in a 4 ^{0}C refrigerator for 16 hours in order to lower the reaction kinetics in cases were the turn-overs were undesirable, such as in blunt-end ligations. After ligation, reactions were heat-shocked in a 65 ^{0}C heat-block for 10 minutes in order to kill the activity of the ligase, and subsequently transformed into competent cells.

2.2.3- Gel Electrophoresis and DNA Extraction

Screening of ligation products and separation of different sizes DNA fragments in restriction digests was achieved by agarose gel electrophoresis. This technique works on the basis on size and charge of DNA molecules by establishing an electric current around an agarose polymer gel, which promotes migration of charged molecules through the matrix. Molecules will run through the gel at different speeds depending on their size: larger DNA fragments have a bigger surface-area of contact with the agarose matrix, and therefore move at a slower rate, while smaller fragments have a lower probability of getting entangled in the gel matrix. The same principle can be applied to the separation of different geometries. For instance, super-coiled plasmid DNA will migrate faster than its open circularized variant due to the smaller surface area. On the other hand, linearized plasmid DNA, which has been excised once with restriction endonucleases, will exhibited the size expected for that DNA molecule, i.e. a 7.5 kb sized linearized plasmid will migrate to the corresponding size on an agarose gel relative to a ladder control.

Agarose is a polysaccharide polymer made up of sub-units of agarobiose that dissolves in near-boiling water, forming a gel after the cooling process. The gels used in this PhD project were made up of either 100 ml or 50 ml of TBE buffer containing 1% of agarose (w/v). It was important for all the agarose gels to contain the same percentage of agarose, as this percentage is directly related to the size of the resulting gel pores, and consequently to the resolution of the gel.

Gel casting was done by heating up the agarose/ TBE solution to boiling-point for 1 minutes, and subsequently pouring the hot liquid into a plastic mold. The size and number of wells was determined by different well combs available as part of the mold. In addition, to allow for the visualization of the DNA fragments after electrophoresis, 5 to 10 µl of 1% (w/v) ethidium bromide was mixed into the cooling gel sample. After casting, gels were loaded into a electrophoresis tank containing positive and negative electrodes and subsequently immersed in TBE buffer.

Prior to loading on the gels, DNA samples were mixed with a dyed loading buffer, which allowed for the samples to be loaded into the submerged wells and for the monitoring of the diffusion rates of samples during electrophoresis. After sample loading, the gel tank was sealed and an electric current established through connecting the electrodes to a powerPac from BioRad with a set voltage. All the agarose electrophoresis experiments were conducted with a electric current voltage of 120 Volts (V).

After electrophoresis, the bands for the different DNA fragments were visualized using a Sygene Bio-imaging system containing a fluorescence UV transilluminator and a capture camera. Photographs of the different agarose gels were taken for subsequent analysis of the DNA samples.

As mentioned in the previous sections, some ligation steps required the separation of a heterogeneous restricted DNA sample and subsequent gel extraction of the desired fragments prior to the ligation reaction. The QIAGEN gel extraction kits were routinely used for the purification of DNA fragments from agarose gels. This type of kit, like the other kits already described, relies on a the use of a solid phase matrix that binds to DNA upon contact. As with other kits, a accuSpin^{TM} Micro centrifuge was used to apply the centrifugal force required for the flow of samples and wash-buffers through the matrix.

The following gel extraction protocol was used to extract DNA fragments from agarose gels:

- The desired bands were first excised from the agarose gel by using a scalpel and a benchtop UV transilluminator to enable the identification of the

correct bands. Exposure of the agarose gel to the UV was minimized to prevent UV-mediated damage to the DNA.

- The excised bands were subsequently transferred to sterile 1.5 ml eppendorf tubes, which had been pre-weighted. The weight of the bands (expressed in mg) was then calculated by subtracting the weight of the empty tubes to the total weight of the tubes containing the agarose bands.
- 3 volumes of QG buffer were subsequently added to 1 volume of gel (where 100 µl corresponds 100 mg of gel), and the resulting samples were incubated in a heat-block at 50 ^{0}C for 10 minutes in order for the bands to dissolve completely.
- 1 gel volume of Isopropanol was then added to the samples, which were subsequently loaded into the QIAquick spin columns. The columns were centrifuged (1 minute, 17 000 g) to bind the DNA to the solid matrix, and the resulting flow-through was discarded.
- Two washing steps were subsequently performed on the columns, in which 750 µl of PE buffer were added to the columns and retained for 5 minutes in contact with the DNA-bound matrix. The washing buffer was subsequently washed via centrifugation (2 minutes, 17 000 g) and resulting flow-through discarded.
- The DNA was subsequently eluted from the columns, through centrifugation (1 minute, 17 000 g), in a final volume of 20 µl of EB buffer. DNA concentrations were checked through a Nanodrop 2000 spectrophotometer with a wavelength of 260 nm.

2.2.4- Amplification of DNA Fragments using PCR

Polymerase chain reaction (PCR) was used to amplify fragments from low concentration DNA samples, and to add specific endonuclease restriction sites at the ends of genes. PCR works through repeating thermal cycles that allow for short sequence primers to anneal to the sequences to be amplifies and for enzymatic replication to occur. Each cycle involved three distinct steps. In the first step the double stranded (ds) template DNA is melted at 98 ^{0}C to producesingle-stranded (ss) targets for primer annealing. The second step involves decreasing the sample

temperature to the optimal annealing temperature of the primers to the target sequence. Primers are designed complement the boundaries of both forward and reverse strands of the sequence in order to create ds DNA products. In the final step, enzymatic replication from the 3' end of the primers is triggered by using DNA polymerases that are activated at a specific temperature. As this cycle is repeated, the replication product from the previous chapter is used as a template for a new replication cycle, resulting in an exponential increase in the concentration of the original template.

The TC512 gradient thermal cycler (TECNE) was used for all the PCR experiments conducted throughout this PhD project. This instrument allows for the set-up of a temperature gradient across different samples in the same reaction, which was very useful to test for the optimal annealing temperature for a given pair of primers. The general cycle program used was the following:

- Melting step at 98 ^{0}C for 10 seconds.
- Annealing step at tm (annealing temperature) of the primers +/- 5 ^{0}C.
- Replication step at 72 ^{0}C, using the Phusion^{R} High-fidelity DNA.

Polymerase (NEB).

This was repeated for 30 cycles, after which a final replication step was performed at 72^{0}C for 10 minutes. The temperature used for the annealing steps depended on the tm calculated for each primers, which was done using the NEB web calculator (https://www.neb.com/tools-and-resources/interactive-tools/tm-calculator). The Phusion^{R} High-fidelity DNA Polymerase (NEB) was for the replication of the amplified fragments due to the fact that it has a lower error rate than other polymerases in the market.

All the primers used in the PCR experiments were designed to contain a sequence of 20 to 25 nucleotides complementary to the ends of the DNA fragment of interest. In addition, in cases where new restriction sites were introduced to the ends of the template sequence, restriction site overhangs that did not anneal with the sequence were introduced at the 5'-end of the primer. When possible, primers were also designed to have similar tm so that the same annealing temperature would be optimal for both primers. The following sets of primers were used:

- Primers for the pCT20 fragment:

Forward strand (tm= 61 ^{0}C)

5- ACAGACAGGACAAAATCGATTTTTAC- 3

Reverse strand (tm= 61 ^{0}C)

5- AGACATCCAAAAATCCGTATTTTGAT- 3

- Primers for the CHMO gene:

Forward strand (tm= 67 ^{0}C)

5- CCCCTCGAGATGTCACAAAAAATGGATTT- 3

Reverse strand (tm= 76 ^{0}C)

5- GGCTGCAGTTAGGCATTGGCAGGTTGCT- 3

- Primers for the GenR gene:

Forward strand (tm= 71 ^{0}C)

5- CTCGAGATGTTACGCAGCAGCAACG- 3

Reverse strand (tm= 74 ^{0}C)

5- GCATGCTTAGG TGGCGGTACT TGGGT- 3

PCR samples were prepared by mixing 1 µl of template DNA with a reaction mix containing the Phusion^{R} High-fidelity DNA polymerase, the corresponding reaction buffer, the set of primers for the amplification process, and the free deoxyribonucleotides. The resulting DNA mix was loaded unto 250 µl sterile eppendorf prior to

PCR. For a 50 µl reaction, the following volumes of the different components were added:

- 1 µl of DNA template.
- 1 µl of 10 mM (milliMolar) dNTPs.
- 2.5 µl of each primer (to a final concentration of 0.5 µM)
- 0.5 µl of Phusion® High Fidelity Polymerase
- 10 µl of High Fidelity buffer (5X)(NEB)
- 32.5 µl sterile H_2O

The different components were thoroughly mixed to ensure that the reaction sample was homogenous. The different primers were provided by Eurofins MWG Operon.

2.2.5- *In-silico* Design of Promoters for Heterologous Protein Expression.

The in-silico design of the promoters used for the expression of heterologous proteins in *S. carnosus* was done through several cycles of criteria formulation, data gathering and processing, construction of designs, and re-evaluation of these designs after experimental testing.

In a first stage of each design cycle, a set of criteria were formulated for the nature of the promoter that would be advantageous to the expression of heterologous proteins in *S. carnosus*. These criteria include: the type of expression system, constitutive or regulated; the restriction map of the promoter region; the source DNA for the promoter; the expression strength of the promoter. The formulation of the desirable criteria were not dependent on what prior knowledge existed on each of the criteria, but rather used as a set of broad directives to guide the promoter screening and design strategy.

The second stage in this process was data acquisition and processing. This entailed the use of several online databases for the acquisition of raw DNA libraries, which would then be processed and filtered for promoter sequences using several DNA manipulation and prediction software. We used the genome sequence of *S. carnosus* as the main source for the construction of the DNA

libraries, which is readily accessible from the NCBI database (http://www.ncbi.nlm.nih.gov/genome/?term=Staphylococcus%20carnosus) . The use of the staphylococcal genome as a source for promoters, as opposed to relying on promoter databases such as the Registry of Standard Biological Parts, seemed more desirable for several key reasons: the promoter databases readily available for use focus on promoters for more conventional bacterial systems like *E. coli* and *S. subtilis* , which have been previously shown to have limited use in Gram-positive staphylococcal systems; the selection of promoter sequences from the staphylococcal genome ensures that the latter are inherently adapted to work in *S. carnosus* as a biocatalytic host. However, there are also inherent risks with the choice not to operate with established databases. One of these risks is that, apart from a couple of promoters already characterized in the literature, there is very little background knowledge about the promoter systems in staphylococcal species, and therefore the uncertainty of selecting promoters with the desired traits is higher.

In the first round of promoter design, a DNA library were constructed from the genome of *S. carnosus* on the basis of the expression of the genes downstream from the selected sequences. Our approach was to only include sequences preceding genes that exhibited an high level of overall expression in the native organism. This decision was taken under the assumption that, while there is not a linear correlation between level of protein expression and transcription of the corresponding gene, it is more likely that an highly level of protein expression is correlated to a high level of gene expression. Subsequently, these genes are more likely to be under the influence of strong promoters.

Since most of the data on the proteomics of staphylococci has not been done on *S. carnosus*, we used the statistical web-server tool CAIcal (Puigbò et al. 2008) for the prediction of highly expressed genes on the basis of their codon composition. The analysis done by CAIcal is based upon the codon adaptation index (CAI), first introduced by Sharp and Li (1987) as a measurement of codon bias of a gene relative to a reference set. CAI quantifies the similarities between the codon usage in a gene and that of the reference set, and can be

mathematically be expressed as the geometric mean of the codon biases for each codon in a gene, following equation 2:

$$CAI = exp\left(\frac{1}{L}\sum_{L=1}^{L} log\ (w(l))\right) \quad (2)$$

Where L is the total number of codons in the gene, and w is the codon bias of each of the codons, which is mathematical expressed as the ratio between the frequency of that codon and the frequency of the optimal codon for the synonymous amino acid (equation 3), which in turn is determined from the reference set:

$$w_i = \frac{x_i}{x_{imax}} \quad (3)$$

The codon frequencies for each codon in a set of genes were also calculated using CAIcal, while the codon usage frequencies of the reference set were taken from the Codon Usage Database (http://www.kazusa.or.jp/codon/)

The different promoters were designed by using several DNA manipulation software. The software SerialCloner was the main tool used for the selection of manipulation of DNA sequences, sequence alignments and analysis of the restriction maps. This software also allowed for the simulation of the ligation steps required for the insertion of each of the designed promoters into the vector construct. Other web-based tools were also used to facilitate the manipulation of the DNA sequences: DNA Massager (http://www.attotron.com/cybertory/analysis/seqMassager.htm) was used to easily reverse-complement sequences, while the NEBcutter (http://tools.neb.com/NEBcutter2/) was used as a restriction site database and as a tool to check compatibility between different restriction sites. In the first round of promoter design, sequences were lifted directly from the *S. carnosus* genome and ordered commercially as synthetic genes, using the provider DNA 2.0. For the subsequent runs of promoter design experiments, we took a modular approach to the design strategy, by independently searching for the different components of the promoter: -35 and -10 transcription elements, ribosomal binding site (RBS), and regulatory elements. These promoters were then physically constructed through

the assembly of short overlapping oligos, which in turn had been provided by Eurofins.

The last stages of the *in-silico* design strategy were the implementation of the designs into a common vector construct, in which the biocatalyst or reporter gene was placed under the expression of the design promoter, and validation of the design through analysis of the experimental data with the resulting ligation. Depending on the experimental results in the protein expression assays, the promoter design was either validated as an efficient expression system, or re-evaluated through a second cycle of *in-silico* design strategy.

2.2.6- Production of Competent Cells

E. coli TOP10 made chemically competent by using a protocol previously devised by Pope et al. (Pope & Kent 1996). Following this protocol, 1 ml of 5 ml overnight cultures was transferred to 250 ml baffled shake flasks containing 50 ml of NB2 medium. The resulting inoculates were grown in Inova 4330 shaking incubators (New Brunswick) at 37 ^{0}C and 250 rpm, until the optical density (OD) of cultures reached 0.7. Cultures were then transferred into Falcon tubes, and the biomass collected by centrifugation (centrifuge 5810 R, 5 min, 16 639 g, 4 ^{0}C). The pellets were subsequently resuspended in half volume of ice-cold 0.1 M $CaCl_2$, and stored in ice for 30 minutes. After this period, cells were re-pelleted by centrifugation (centrifuge 5810 R, 5 min , 1 238 g, 4 ^{0}C), and resuspended in 2.5 ml of 0.1 M $CaCl_2$, which was used directly in the transformation protocols or fractionated into 100 µl aliquots containing 10 % (w/v) glycerol and stored at -70 ^{0}C. It was important to maintain the temperature of bacterial cultures below room temperature in order to prevent potentially harmful stress responses from the cells during the incubation with $CaCl_2$. In addition, the re-suspension of cultures was done gently in order to minimize shear damage.

Transformation of *E. coli* chemically competent cells was done by mixing 5 to 10 µl of DNA with 100 µl cell aliquots, and subsequently storing the mix on ice

for 30 minutes. Samples were subsequently heat-shocked in a 42 ^{0}C water-bath for 30 seconds and then quickly stored on ice for another 2 minutes. After heat-shock treatment, 0.5 ml of NB2 medium was added to samples, which were subsequently incubated for 1 hour at 37 ^{0}C and 250 rpm. The resulting cultures were plated unto 2% (w/v) agar plates containing 100 µg/ml of ampicillin or 15 µg/ml gentamicin, depending on selective gene present on the plasmid DNA. Plates were incubated for 24 hours at 37 ^{0}C.

For transformation of *S. carnosus* cultures, electrocompetent cells were produced by following an optimized version of the protocol developed by Löfblom *et al*. (Löfblom et al. 2007). In the optimized protocol, staphylococcal aliquots from 5 ml overnight cultures grown on B2 medium were transferred to falcon tubes containing 5 ml of B2 medium until the final OD_{600} of the solution reached 0.5. The resulting inoculate were grown for 2 hours at 37 ^{0}C and 250 rpm, subsequently incubated for 15 minutes in ice to stop bacterial growth, and the biomass collected by centrifugation (5 min, 16 639 rpm, 4 ^{0}C). The cultures are subsequently washed twice with 2.5 ml and 250 µl of ice-cold water by centrifugation (10 min, 16 639 rpm, 4 ^{0}C). The washed pellets were resuspended in a final volume of 50 µl of a 10% (w/v) glycerol solution and used directly for transformation.

An MicroPulser^{TM} electroporator (BioRad) was used for the transformation of electrocompetent *S. carnosus*. Aliquots of electocompetent cells were mixed with 10 µl of DNA and incubated at room temperature for 30 minutes. The mixed samples were subsequently transferred to 0.2 cm (centimeter) electroporation cuvettes (BioRad), and electroporated using a single 1.1 ms (milliseconds) pulse with a voltage of 2.1 Kv (kilo-volts) cm^{-1} . Electroporated samples were subsequently resuspended in 0.5 ml of B2 medium, and incubated for 2 hours at at 37 ^{0}C and 250 rpm. The resulting cultures were plates unto 2% (w/v) agar plates containing 15 µg/ml of chloramphenicol or gentamicin, depending on the selective gene present in the DNA. Plates were incubated for 48 hours at 37 ^{0}C.

An alternative method that did not involve the pre-production of competent cells was also used. This protocol was previously developed by Lin *et al.* (L. Lin et al. 2010), and involved mixing 500 µl aliquots from 5ml overnight cultures grown in B2 medium with 10 µl of DNA and sonicating the resulting sample at an high

amplitude. A Soniprep 150 Plus sonicators (MSE, 23 KHz (Kilohertz) was used for the sonication process, at an amplitude of 20 microns for 45 seconds. An equal volume of B2 medium was subsequently added to the samples, which were incubated for 4 hours at 37 ^{0}C and 250 rpm. The resulting cultures were spread unto agar plates containing 15 µg/ml of chloramphenicol or gentamicin. Plates were subsequently incubated for 48 hours at 37 ^{0}C.

2.2.7- Plasmid Stability Assays

The plasmid stability assays were conducted on both plasmid containing *E. coli* and *S. carnosus* cultures using the same protocol. Cultures were first grown overnight in 5 ml of NB2 medium containing the corresponding antibiotic, after which 100 µl aliquots of the cultures were inoculated into falcon tubes containing 5 ml of fresh NB2 medium. The inoculates were then grown for 72 hours in the absence of antibiotic in an Inova 4330 shaking incubator at 37 ^{0}C and 250 rpm, with 100 µl of inoculate being transferred to falcon tubes containing 5 ml of fresh medium after every 24 hours.

Samples were taken twice during each 24 hours period, and subsequently plated unto both antibiotic containing 2% (w/v) agar plates and plates without antibiotic. Prior to being spread on the plates, samples were diluted so that the number of colonies growing on the plates did not exceed 10^2 magnitudes. Plates were incubated overnight at 37 ^{0}C, and results from the assay were expressed as percentage of surviving plasmid-containing (pl) colonies, which could be calculated from the fraction between number of colonies grown in antibiotic selective plates and in plates without antibiotic (equation 4):

$$\%\, pl\ colonies = \frac{N.of\ colonies\ in\ antibiotic\ plates}{N.of\ colonies\ in\ non-selective\ plates} * 100 \qquad (4)$$

2.3- Biocatalytic Techniques:

2.3.1- General Growth Kinetics

Most studies performed on *E. coli* and *S. carnosus* strains were conducted in non-baffled conical shake flasks of working volume of 1/5 of the total flask volume. Shake flasks were invariably incubated using a Inova 4330 shaking incubator, with conditions being set at a constant temperature of 37 ^{0}C and shaking speed of 250 rpm, unless specifically specified. It is important that conditions were maintained constant between different batches and bacterial strains since it facilitated comparison of growth, expression and biocatalytic performance between the different cultures. During incubation, the optical density (OD) of growing bacterial cultures was measured routinely by loading a 1 ml sample unto a cuvette and measuring the OD in spectrophotometer set-up to record the absorbance of samples at a wavelength of 600 nm.

The OD measurements of growing cultures were used to plot growth curves, and from the latter determine simple kinetics of bacterial strains under growing conditions. We mainly focused on the lag-time, maximum growth, and the maximum growth rate. The lag-time (ΔT_{lag}) specifies the time between start of incubation and beginning of exponential phase, and is calculated as the difference between the observed time (T_{obs}) that takes for cultures to reach a specific density (x_i) from the start of incubation (x_0) and the time it would take if the lag phase was absent (T_{pred}), i.e. if bacterial cultures grew in exponential phase from the start of incubation (equation 5):

$$\Delta T_{lag}(h) = T_{obs} - T_{pred} \Leftrightarrow \Delta T_{lag} = T_{obs} - \frac{log_2 x_i - log_2 x_0}{R} \quad (5)$$

Where R is the exponential growth rate.

The maximum growth (G_{max}) refers to the point at which the bacterial cultures reach the highest density (x_{max}), and can be calculated using equation 6:

$$G_{max} = x_{max} - x_0 \quad (6)$$

Finally, the maximum bacterial growth rate (μ_{max}) refers to the maximum rate of cell division, which also corresponds to the exponential growth rate (R), and expressed using the following equation (equation 7):

$$R\ (h^{-1}) = \frac{\log_2 x_2 - \log_2 x_1}{T_2 - T_1} \quad (7)$$

Where x_2 and x_1 refer to the densities at different points of the exponential phase, with the corresponding times T_1 and T_2. All the different kinetic factors were determined following the Monod model for bacterial growth kinetics (Monod 2003).

For the specific cases where different containers where used for bacterial growth, i.e. overnight incubation of colonies and rapid DNA screening via a 96-well plate, no kinetic factors were calculated. This was due to the fact that the experiments performed under the latter conditions were qualitative, and therefore it was not essential to acquire quantitative data on the general kinetics of bacterial growth.

2.3.2- Measurement of Dry Cell Weight (DCW)

The dry cell weight (DCW) of cell cultures was determined from the calibration of dry mass as a function of optical density (OD) of bacterial cultures. This calibration was performed by drying bacterial cultures with different OD on 0.22 micron nylon filters from Millipore, which had in turn been previously dried for 5 days in a 50 ^{0}C incubator. The long drying period was performed to stabilize the weight of the filter and to ensure that the latter did not vary throughout the cell drying protocol.

Cells were inoculated overnight in 5 ml of growth medium contained within a shake flask and later transferred into a 500 ml shake flask containing 100 ml of growth medium with the respective selective antibiotic. Cultures were incubated in a shaking incubator for 24 hours at a temperature of 37 ^{0}C and shaking speed of 250 rpm, with 5 ml samples being taken at different cell densities. The cultures from these samples were subsequently drained onto the filters by applying a

vacuum to the output of the filter. After successful filtration, filters were stored in non-sealed pretty dishes and incubated in the 50 ^{0}C incubator until the dry weight of the bacterial cultures stabilized. As a control, clean medium was also filtered unto the same type of filters and dried in parallel to the bacterial cultures. This control was necessary to measure the contribution of medium salts that are not evaporated during the desiccation stages to the overall dry weight of cells cultures. Both controls and working samples were analyzed in triplicates, in order to calculate the degree of confidence of the results.

The final bacterial dry weight for each corresponding OD_{600} was calculated as (equation 8):

$$Final\ DCW\ (g) = Total\ weight - (weight\ of\ filter + weight\ of\ control) \quad (8)$$

The final dry weight of cultures can be converted to dry cell weight concentration (C_{DCW}) by using equation 9:

$$C_{DCW}\ (g.L^{-1}) = \frac{Final\ DCW}{Volume\ of\ culture} \quad (9)$$

The final step was the calibration of C_{DCW} against its corresponding OD_{600}, resulting in a linear correlation, which could be used to calculate the dry cell weight of a culture at given OD_{600}.

For *E. coli* cultures, this linear correlation was expressed using equation 10:

$$C_{DCW}\ (g.L^{-1}) = 0.25 * OD_{600} \quad (10)$$

While for S. carnosus it was expressed using equation 11:

$$C_{DCW}(g.L^{-1}) = 0.25 * OD_{600} \quad (11)$$

2.3.3- SDS-PAGE Analysis of Protein Fractions

Protein extracts from bacterial cultures were examined via SDS polyacrylamide gel electrophoresis (PAGE). This technique is based upon the stimulated migration and fractionation of charged macromolecules through a polyacrylamide gel by applied an electric field across the length of the gel. In a native PAGE, migration is largely determined by the electrophoretic mobility of the macromolecules, i.e. the size, conformation and charge of different proteins.

In SDS-PAGE, sodium dodecyl sulfate (SDS) is added to the samples prior to loading on the gel in order to linearize proteins by disrupting non-covalent bonds within the protein and to impart a negative charge to the latter, which is evenly distributed per unit of mass. In this case, migration of a group of macromolecules through the gel is mainly dependent upon the sizes of the different proteins, as SDS negates both shape and charge of the proteins. As the electric field is established across the gel, samples migrate towards the positive electrode (anode), and separation of the different proteins contained within the latter occurs due to the different speeds at which differently sized protein migrate towards the anode.

Samples used for SDS-PAGE analysis of protein extracts were prepared from 50 ml bacterial cultures grown in 250 ml conical shake flasks with selective antibiotic, on Inova 4330 shaking incubators set up with a constant temperature of 37 ^{0}C and rotation speed of 250 rpm. After collection of the cultures via centrifugation (16 639 rpm, 5 min, 4 ^{0}C), pellets were resuspended in 5 ml of either 50 mM phosphate buffer saline (PBS) (pH 7.4) or 50 mM Tris-HCl buffer (pH 7). The resuspended cultures were subsequently lysed through the use of a French® Pressure Cell Press (Thermo) with a 30 ml maximum capacity stainless steel pressure cylinder and a 1 inch Piston, which had been stored beforehand at 4 ^{0}C to ensure that the temperature of the samples did not rise above the threshold for protein stability during the homogenization process. Samples were also stored on

ice throughout every step of the protocol after collection to minimize any compromise to the stability of proteins within the cells.

During French press homogenization, liquid samples are loaded unto a pressure cylinder and compressed by the use of an external hydraulic pump that drives a piston into the cylinder, building up a pressurized environment inside the cylinder. As a valve is slowly opened at the outlet of the cylinder, cells suffer severe shear stress from decompression, resulting in the disruption of cell membrane.

To ensure efficient cell lysis, *E. coli* samples were submitted to two round of homogenization at a final pressure of 1800 bar, while *S. carnosus* cultures were submitted to three round under the same conditions.

The soluble and insoluble fractions of the lysates were subsequently separated using a bench top centrifuge (16 639 g, 5 min, 4 ^{0}C) and aliquots were taken from each fraction individually for loading unto the SDS-PAGE polyacrylamide gel. Aliquots from the soluble fraction were taken directly from the supernatant of lysates and mixed with protein loading buffer, while the insoluble fraction was first resuspended in an equal volume of PBS or Tris-HCl buffer before mixing with the loading buffer. In turn, the protein loading buffer was aliquoted from a 2X stock solution containing 0.8 g of SDS, 0.4 ml of β-mercaptoethanol, 8 mg of Coomassie blue G-250, and 4 ml of 100% (w/v) glycerol for every 10 ml of solution.

After mixing of the loading buffer with the supernatant and pellet extracts, the samples were heated to 95 ^{0}C in order to facilitate the linearization of protein structures by SDS and β-mercaptoethanol, and to melt down any cell aggregates that might otherwise hinder the loading of the samples into the polyacrylamide gel. The samples were subsequently spun down in a benchtop centrifuge at maximum speed for 1 minute to separate any remaining aggregates, and loaded onto the polyacrylamide gel.

The polyacrylamide gel was prepared in distinct phases: the running phase and staking phase (Figure 13). The running phase is the first to be settled onto a glass cassette, and as the name implies it is the phase where the samples migrate and are fractioned into distinct bands according to size. It was prepared through a mixture of 2 different liquid buffers supplied by National Diagnostics, ProtoGel 30% and Resolving Buffer (4X), and distilled water.

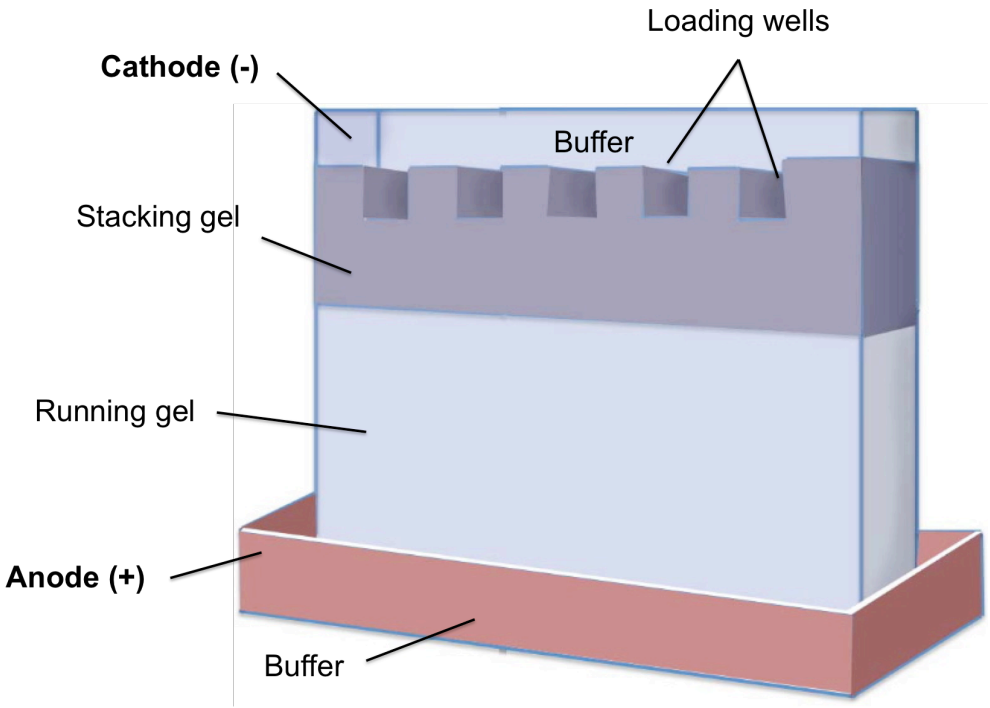


Figure 13- Diagram representing the different layers of an SDS-PAGE gel (http://www.siumed.edu/).

The Protogel 30% is the component of the mix that contains the polymerizing acrylamide monomer and bisacrylamide, and depending on the desired final percentage of acrylamide on the gel, different volumes of the buffer are added to different dilution volumes of distilled water. By contrast the volume of resolving buffer remains constant regardless of the final gel percentage. The percentage of acrylamide in the resolving phase determines the size of the pores through which the linearized proteins migrate through, and consequently determines the overall resolution of the gel. Therefore, it was very important to maintain the acrylamide percentage of the constant throughout the multiple SDS-PAGE assays conducted in the protein studies in order to ensure that results from different assays were comparable. To achieve a resolving phase with a percentage of 12%, we mixed 40 ml of ProtoGel 30% with 33.9 ml of distilled water and 25 ml of resolving buffer. Acrylamide is capable of spontaneous autopolymerization upon hydration, forming long single-chain polymers. However,

this form of polymerization is slow, and results in a viscous solution rather than a solid phase due to the fact that the polymers are not cross-linked. Therefore, prior to loading the mixture unto the gel cassette, a small volume of the chemical initiators ammonium persulfate (APS) (1% v/v) and TEMED (0.1% v/v) were added to speed up polymerization and to trigger cross-linking between acrylamide polymers and bisacrylamide molecules present in the ProtoGel 30%. After addition of initiators, the resolving gel mixture was poured into the gel glass cassette, which was set up vertically to allow the gel to settle through gravitational pull.

The stacking phase was prepared in a similar fashion, by mixing 1.3 ml of ProtoGel 30% with 6.1 ml of distilled water and 2.5 ml of stacking buffer (4X). In this case the volumes of distilled water and ProtoGel 30% are not variable, since the percentage of acrylamide in the stacking gel does not contribute to the overall resolution of the results. TEMED (0.01 ml) and APS (0.05ml) were also added for reasons stated previously. Once the resolving phase was solidified, the polymerizing stacking phase was poured on top, and sample wells were molded through the use of a flat comb with broad indentations, which was subsequently removed once the stacking phase was completely settled.

The resulting polyacrylamide gel was loaded into a transparent plastic chamber containing the electrode circuit, and immersed in a SDS-PAGE running buffer that had been previously diluted from a 10X stock solution. Pre-treated pellet and supernatant samples were subsequently loaded into the immersed wells, and electrophoresis is started by applying an electric current via a powerPac from BioRad that had been connected to the electrophoresis apparatus. In all the SDS-PAGE assays conducted throughout this project, the electric current was set to amplitude of 35 mAmps and a electrophoresis running time of 50 minutes.

The last step in the SDS-PAGE protocol is the staining and de-staining of the protein gel, which is necessary for the visualization of protein bands on the gel. The staining solution we used was a Coomassie Brilliant Blue based stain, which binds non-specifically to proteins. In the staining protocol, we added varying volumes of the staining solution to the protein gel, which had been removed from the glass cassette after electrophoresis and placed in shallow plastic container. Staining was achieved by heating the immersed gel in the microwave at maximum power for 1 minute, or alternatively by gently mixing the immersed gel for 30

minutes on a revolving table. De-staining was achieved through several washing steps in which the gel was re-immersed in a variable volumes of distilled water and subsequently heated in the microwave at maximum power for 5 minutes.

The gels were visualized by using a light-box contained within a sealed chamber with a capture camera on top.

2.3.4- NADPH Oxidation Assays

The activity of intracellularly expressed biocatalyst was readily detected in the lysates of bacteria hosts by measuring the oxidation rates of NADPH. During the conversion or cyclohexanone into ε-caprolactone, CHMO uses NADPH as a source of electrons, and therefore enzymatic activity can be correlated to the amount of NADPH oxidation over time. Since NAPDH absorbs at 340 nm, the rates of oxidation can be measured as the decay in absorbance of samples containing lysates and reaction buffer at 340 nm as a function of time.

The bacterial lysates were prepared by growing overnight inoculates in 250 ml shake flasks containing 50 ml of growth media with the corresponding selective antibiotic. After 8 hours of growth in an INOVA 4330 shaking incubator at 37 ^{o}C and 250 rpm, cells were harvested by centrifugation at (5810 R, 16 639 g for 5 minutes, 4 ^{0}C), using a bench top centrifuge 5810 from Eppendorf with a rotor for 50 ml Falcon tubes. Before cell harvesting, the optical density of the cultures was measured using a spectrophotometer at 600 nm in order to calculate the dry cell weight of the pellets. The cell pellets were then resuspended in 5 ml of 50 mM Tris-HCl (pH 9) buffer, and subsequently lysed via mechanical homogenization using a French press. The same method was used for the mechanical lysis of both *S. carnosus* and *E. coli* cultures, with cells being loaded onto the pressure cylinder and submitted to 1800 bar twice. After homogenization, soluble lysate was separated from the insoluble cell debris via centrifugation (16 639 g, 5 min, 4 ^{0}C), and subsequently aliquoted into 0.2 ml samples that were used directly in the NADPH oxidation assays. All cultured samples were stored in ice throughout the protocol to ensure that the cells were in a dormant state and no unwanted enzymatic activity occurred during cell lysis. Storing samples at low temperatures

also increased the stability of soluble proteins that were released to the supernatant upon lysis.

The NADPH oxidation assays were performed following an protocol established by Doig *et al.* (2002). Lysate aliquots were loaded into 1 ml cuvettes containing 0.8 ml of reaction buffer, which had been prepared beforehand and stored at 4 ^{0}C. The reaction buffer was composed of 50 mM Tris- HCl (pH 9), 0.161 mM of NADPH, and 7.14 g/L of BSA (Bovine Serum Albumin). The cuvettes containing the mixture were subsequently loaded into a spectrophotometer with temperature control set to 30 ^{0}C, and the rates of oxidation of NADPH were measured at an absorbance of 340 nm for 2 minutes in the absence of the biocatalytic substrate cyclohexanone. This rate corresponds to the basal activity of the proteins in the lysate, as there is always a percentage of intracellular proteins that oxidize NADPH. After the first 2 minutes, a varying concentration of cyclohexanone (2- 8 mM) is added to the reaction cuvettes from a stock solution of 1 M, and the NADPH oxidation is measured for another 2 minutes. The resulting oxidation rates corresponded to the sum of activities from the intracellular active contents of the cell and the biocatalytic activity of CHMO in the conversion of cyclohexanone to caprolactone. From these results we were able to calculate the specific activity of CHMO in the lysates.

Unfortunately, we were unable to find a protocol that could give an accurate measurement of the amount of CHMO in the lysates: We were unable to purify active protein from cell extracts; Bradford assays would only give a rough estimate and lacked the sensitivity to detect small variations in biocatalyst concentration, which considering the difficulties in pinpointing CHMO overexpression in SDS-PAGE gels, was not very useful as a measurement of precise amounts of the protein. Therefore, the specific enzymatic activity was expressed as amount of NADPH oxidized per minute per gram of dried cell weight (DCW).

Specific enzymatic activity was calculated using the following steps :

- The NADPH oxidation rate (ω_{basal}) (equation 12) from the lysate basal activity was first subtracted from the overall NADPH oxidation rate (ω_{total}) (equation 13) after the addition of the biocatalytic substrate, in order to obtain oxidation rate resulting from the biocatalytic activity of CHMO (ω_{CHMO}) (equation 14):

$$\omega_{basal}(abs.min^{-1}) = |Abs_{t=2} - Abs_{t=0}| \quad (12)$$

$$\omega_{total}\,(abs.min^{-1}) = |Abs_{t=4} - Abs_{t=2}| \quad (13)$$

$$\omega_{CHMO}(abs.min^{-1}) = \omega_{total} - \omega_{basal} \quad (14)$$

- The values of absorbance can be translated to amount of NADPH by using the extension coefficient (**ε**) of NADPH at a wavelength of 340 nm (equation 15):

$$C\ (mol.L^{-1}) = \frac{Abs}{l * \varepsilon} \quad (15)$$

C = concentration ($mol.L^{-1}$)

ε = extension coefficient = 6.23 x 10^3 $L.mol^{-1}.cm^{-1}$ for NADPH at 340 nm

l = light path length (1 cm)

The concentration of NADPH can be in turn converted to the amount of NADPH (**m**) present in the 1ml reaction sample (equation 16):

$$m\ (mol) = C * 10^{-3} \quad (16)$$

- Finally, the specific enzymatic rate of CHMO (μ_{CHMO}) in the lysates was expressed using the following equation 17:

$$\mu_{CHMO}\ (mol.min^{-1}.g(DCW)^{-1}) = \frac{\omega_{CHMO}}{G} \quad (17)$$

Where **G** is the dry cell weight of the cultures from which the lysates were extracted, and can be calculated for each 0.2 ml aliquot using equation 18:

$$G\ (g(DCW)) = \frac{(C_{cell} * Volume\ of\ cultures\ (50ml)) * 0.2}{5} \quad (18)$$

Where the constant factors 0.2 and 5 correspond to the concentration of volume that occurs as cell cultures are pelleted from the 50 ml volume and resuspended in 5 ml of Tris-HCl buffer. Both factors are in the equation expressed as milliliters (ml), and therefore required prior conversion into litters (L) in order to correspond to the units of the concentration of cells grown in the shake flasks (C_{cell}), which in turn was calculated using equation 19:

$$C_{cell}\ (g(DCW).L^{-1}) = OD * C_{OD=1} \quad (19)$$

$C_{OD=1}$ = concentration of cells when the optical density (OD) is 1, which had been determined previously.

Alternately, the units of C_{cell} could be converted to grams (g) per milliliter (ml), which in the end was preferred in order to simplify equation 9.

2.3.5- Whole-Cell and Resting-Cell Biocatalysis Assays

In order to access the efficiency of *E. coli* and *S. carnosus* strains to work as efficient biocatalysis host, whole-cell biocatalytic assays were performed on both actively growing bacterial cultures and cultures that have been resuspended in non-growing medium and subsequently maintained in the stationary phase.

The reasoning behind performing biocatalysis on actively growing cells was to access the overall performance of CHMO biocatalysis when the cell metabolic machinery is consuming oxygen, which is a competing resource. Moreover, the analysis of biocatalytic rates during bacterial growth would allow us to access the optimal OD for transferring the cell cultures to non-growing medium, i.e. the OD at which the bacterial cultures exhibited the highest activity rates. These assays were performed by transferring an overnight bacterial inoculate into 250 ml Falcon tube

containing 50 ml of growing media, the corresponding selective antibiotic, and a non-inhibitory concentration of cyclohexanone, which ranged from 10 mM to 20 mM (taken from a stock solution of 1 M cyclohexanone). Cultures were grown for 24 hours in a INOVA 4330 shaking incubator at 37 ^{0}C and 250 rpm, with samples being taken at different time points for the subsequent measurement of cyclohexanone and caprolactone content via gas chromatography (GC), and for measurements of optical density (OD). Samples were taken in triplicates in order to determine the confidence range of the results.

For the resting-state biocatalytic assays, cells were grown as previously described in the absence of biocatalytic substrate until the end of exponential phase, or at the point at which the biocatalytic rates are optimal (determined from the actively growing cell experiments), which for most bacterial cultures corresponded to a bacterial growth of 8 hours. Cells where subsequently harvested via centrifugation (16 639 g, 5 min, 4 ^{0}C) and resuspended in 5 ml of phosphate buffer saline (pH 7.4) containing 10 g/L of glycerol, as a carbon source to allow for basal cellular activity, and 10-20 mM of cyclohexanone. Incubation of cultures under these conditions was performed in a shaking incubator at 37 ^{0}C and 250 rpm during Samples were taken at specific time points for subsequent GC analysis of substrate/product contents.

In both the active-growing and resting-state biocatalytic assays, the whole-cell biocatalytic activity rate ($\boldsymbol{\mu_{cell}}$) was calculated using a modified version of equation 8 (equation 20):

$$\mu_{cell}\ (mol.h^{-1}.g(DCW)^{-1}) = \frac{[product/substrate]*vol.\ of\ culture}{G*time} \quad (20)$$

Where **G** is calculated using a modified version 9, in which the constant factor 0.2 and 5 are removed. This is due to the fact that the **G** in both active and resting-state assays is the same, and therefore no correction of the value due to change in final volume is required. Despite the concentration step performed on cell cultures prior to resting-state biocatalysis, the final mass of cultures does not change with the reduction of volume.

2.3.6- GC Analysis of Whole-Cell Biocatalytic Samples

The detection of cyclohexanone and ε-caprolactone in samples taken from the biocatalytic assays was done on a autosystem XL-2 gas chromatograph (Perkin- Elmer). Gas chromatography (GC) is a technique for analytical chemical studies that relies on the separation of different vaporized compounds as they are carried through a liquid stationary phase with a specific length and polarity. The GC chromatograph set-up is generally composed of three main components: an inlet, where the samples are introduced through a syringe and vaporized. The vaporized analytes are subsequently transported by a gas stream, designated the mobile phase, through a column that contains a liquid stationary phase. As the samples are carried through this column, the various analyte components will be retained in the stationary phase to differing degrees. The interaction of the different compounds with the stationary phase causes each compound to be eluted through the column at different times. At the output of the column, a detector is used to measure the amount of analytes as they exit the system.

The different retention times of compounds, i.e. the time in which they are retained within the stationary phase, are what determines the analytical power of the GC system, as the identification of specific compounds is dependent on the ability to discern the retention times of different analytes. In turn, retention time is dependent on two major parameters: the polarity of the stationary phase, and the temperature of the oven in which the column is contained. The polarity of the stationary phase determines the level of interaction between compounds and the column, while the temperature of the controlled column environment dictates the rate at which different compounds are carried through the column. Similar polarities between the stationary phase and sample analytes result in increased retention time of the latter, thereby slowing the analysis process; on the other hand, increased diffusion rates of samples through the column decrease the residence time of the analytes in the column. The interplay between these factors determines the resolution of the GC system. In addition, column length and diameter also

affect resolution by increasing or decreasing the general residence time of samples in the column.

In the experimental set-up used throughout this PhD project, 1 µl fractions of samples from the whole-cell biocatalytic assays were loaded by an in-built syringe into a splitless injector set at 280 ^{0}C, and subsequently carried through a Heliflex AT-1701 column 30 m (length) X 0.54 mm (diameter). Helium was used as the mobile phase, and set at a constant pressure of 27 Psi (pounds per square inch). The oven temperature was initially set for 50 ^{0}C and maintained constant for 2 minutes after injection of the samples. This temperature was subsequently increased at a rate of 20 ^{0}C min^{-1} until a final temperature of 260 ^{0}C, which was maintained for another 2 minutes. A FID (flame ionization detector) fuelled by hydrogen and air and set at a temperature of 280 ^{0}C was used for the detection of the different biocatalytic compounds, which were displayed as peaks in a 2D spectra as a function of retention time.

Cyclohexanone, which has a boiling temperature of 155.6 ^{0}C, was expected to have a lower retention time than ε-caprolactone, which has a boiling temperature of 253 ^{0}C. Accordingly, in all the GC assays, both compounds could be easily differentiation, with the cyclohexanone peak appearing at a retention time of around 2.7 minutes after injection, and the caprolactone peak appearing 6.7 minutes after injection.

The concentrations of the different compounds were derived from the area of the corresponding peaks, which is linearly proportional to the amount of compound in the injected sample. To do so, calibration of the GC protocol was performed for both cyclohexanone and ε-caprolactone in the beginning of each GC analysis, In these calibration protocols, a range of set concentrations for each compound were loaded into the GC chromatograph, and the peak area generated by each concentration was subsequently measured and plotted against the latter. Figure 14 shows the 2D spectra for the calibration of cyclohexanone, in which an increase in the concentration of the compound resulted in a proportional increase in the area of the peak. Results obtained from the calibration experiment were subsequently used to derive the quantitative relationship between peak area and compound concentration, plotted as calibration curves, from which the

concentration of cyclohexanone and ε-caprolactone in the biocatalytic samples was calculated using equation 21:

$$[compound]\ (mol.L^{-1}) = m * peak\ area \qquad (21)$$

Where **m** is the slope obtained from the calibration curves of cyclohexanone and ε-caprolactone.

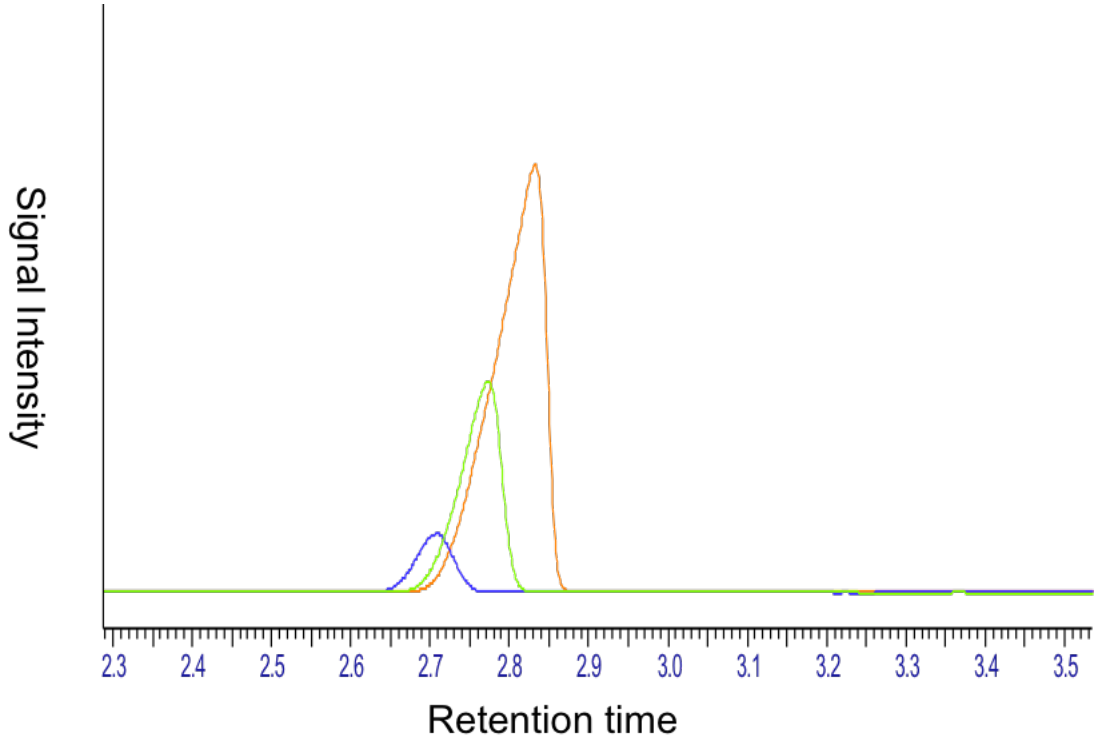


Figure 14- 2 D spectra from a calibration experiment done with three different concentrations of cyclohexanone: 15 mM (blue trace), 50 mM (green trace), and 100 mM (red trace).

Prior to GC analysis, samples taken from whole-cell biocatalytic experiments were transferred to ethyl acetate in order to prevent any potential damage that a aqueous solution would have on the stationary phase of the column. For this purpose a set volume of sample was mixed with an equal volume of ethyl

acetate in a sterile 1.5 ml Eppendorf tube. The resulting solution was incubated in a Confort thermomixer (Eppendorf) for a range of times at room temperature and 900 rpm. After incubation, samples were centrifuged for 10 seconds at max speed, and the resulting top layer containing the ethyl acetate solvent was subsequently transferred into a sealed chromatography tube, which prevented evaporation of the solvent and any transferred volatile compounds. The optimum time of incubation for achieving maximum transfer of cyclohexanone and ε-caprolactone from the aqueous Tris-HCl buffer to the organic solvent was determined by running ethyl acetate samples incubated for different times in the GC chromatograph. Figure 15 shows the results from one such experiment conducted, in which samples with the same concentration of cyclohexanone were incubated with ethyl acetate for a range of different times. From the results, it could be determined that the optimum incubation time was 1 hour.

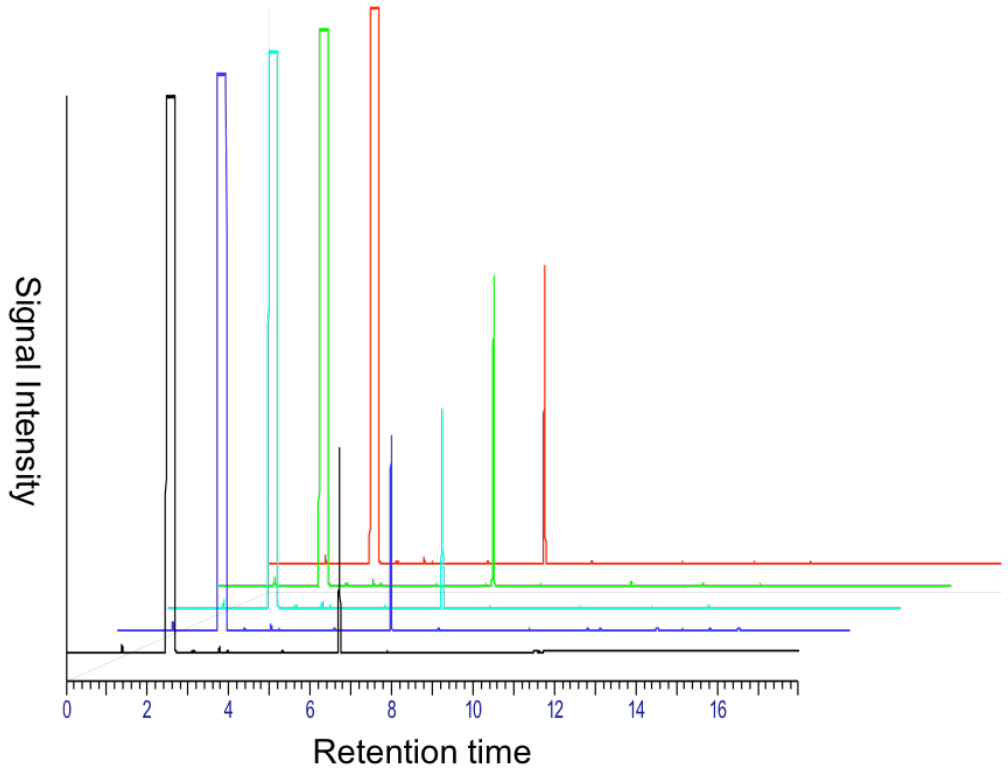


Figure 15- 2D spectra obtained from samples incubated with ethyl acetate for different time frames: 5 minutes (black trace), 15 minutes (blue trace), 30 minutes (light blue trace), 1 hour (green trace), and 2 hours (red trace).

2.3.7- Substrate/Product Tolerance Assays

The tolerance of the different bacterial strains to the substrate cyclohexanone, the product ε-caprolactone and the solvent cyclohexane, was tested by growing cell cultures in the present of these compounds. For this purpose, several different protocols were conducted, in which the compounds were added at different time points of bacterial growing.

For the cyclohexane tolerance assays, 50 mM and 100 mM of the compound were added from a stock solution of 1 M to 250 ml falcon tubes containing 50 ml of NB2 medium with the corresponding selective antibiotic. 1 ml inoculates from 5 ml overnight cultures were subsequently transferred to the shake flasks, and the resulting cultures were incubated for for 10 hours in a Inova 4330 shaking incubator at 37 ^{0}C and 250 rpm. 1 ml samples were taken at specific time point during the incubation for measurements of the cell density, using a spectrophotometer set at a wavelength of 600 nm.

In an alternative protocol, which was used for cyclohexane as well as cyclohexanone and ε-caprolactone, 50 mM and 100 mM of the different samples were added to growing 50 ml shake flask incubations during exponential bacterial growth, between OD_{600} 1 and 2. Samples were also taken from the growing cultures at regular intervals to measure the effect of the different compound concentrations on cell density.

2.3.8- Gentamicin Resistance Assays

Gentamicin resistance assays were conducted on cultures grown in 2 ml v-bottom 96- microwell plates (USA Scientific). Aliquots from over-night cultures were transferred to 96-well plates containing 1 ml of growth medium (NB2 for *E. coli* and B2 for *S. carnosus*) with increasing concentrations of the gentamicin antibiotic. The plates were sealed with alluminium foil and subsequently incubated for 24hours in

a Confort thermomixer set for a temperature of 37 ^{0}C and a shaking speed of 900 rpm. At the end of incubation, bacterial tolerance to the different concentrations of gentamicin was correlated to the final OD_{600} of samples. Figure 16 shows a typical set-up for this experiment.

Alternatively, the response of different bacterial strains was also measured by plating diluted aliquots of over-night grown cultures in 2% agar plates containing increasing antibiotic concentrations, as well as plates without any antibiotic. In this case, results were expressed as % of surviving colonies, which has already been described previously in this chapter (see section 2.2.6).

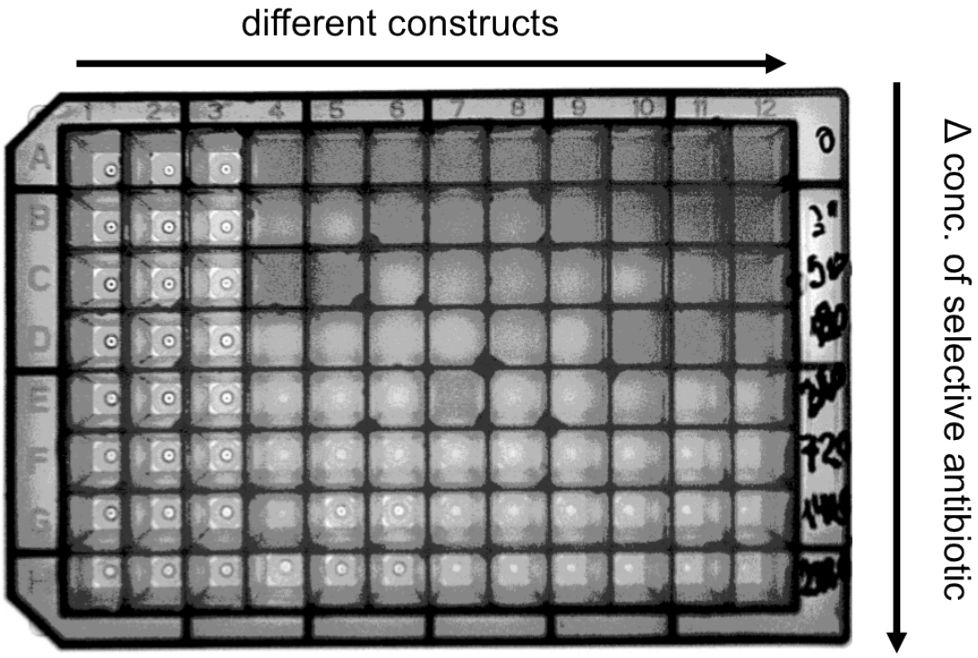


Figure 16- Gentamicin resistance assay performed on 2ml 96-well plates

Chapter 3: Results- Assessment of pQR493 as an Expression Vector for S. carnosus.

3.1- Previous Work on an Expression System for *S. carnosus* and Creation of pQR493

As it was described in the first chapter, some work had already been done previously to validate the use of *S. carnosus* as a host for heterologous protein expression, which resulted in the construction of several expression vectors tailored specifically for this bacterium. The initial idea behind this project was to use one of the expression systems previously constructed and adapt it for expression of CHMO. The chosen expression system was the vector pNW21, which had been used for the expression of the *S. aureus* protein hysA (Williams et al. 2002). This shuttle vector contained elements for replication and selection in both the Gram-negative bacteria E*. coli* and the Gram-positive *S. carnosus*, thus making it very versatile for cloning between the two species. In addition, the vector contained the promoter for the lipase gene from *Staphylococcus hyicus*, which by now had been used extensively as the standard promoter for protein expression in *S. carnosus*. The promoter worked well in Gram-positive strains, but it was shown not to be a very effective promoter for the Gram-negative counterparts (de Vos et al. 1997) (Demleitner & Götz 1994) (Samuelson et al. 1999). Thus, even though the replication and selective elements in the vector allowed for easy genetic exchange between different strains, pNW21 was constructed with the strict purpose of protein over-expression in *S. carnosus*.

In the work that immediately preceded this project, pNW21 was modified to accommodate the protein CHMO. This new construct, designated pQR493, was the primary expression system to be used for intracellular biocatalyst production. However, this system had not been expressed or characterized in *S.carnosus*, and this work needed to be done before proceeding with the main line of enquiry of the project.

3.2- Regressive Analysis of the Restriction Map of pQR493

The early studies conducted on pQR493 did not give very encouraging results. Not only attempts to clone the vector into *S. carnosus* were unsuccessful, but the restriction maps created for the vector did not seem to coincide with the predicted results. Therefore the initial aim for this project was to investigate the genetic integrity of pQR493 and access if it could still be used as an expression system for over-expression of CHMO.

The most direct way to access the integrity of small-scale genetic material is via restriction analysis. Restriction maps outline the restriction endonucleases that can cut a sequence, where they cut it, how often they can cut it, and the distance relationship between the restriction sites. Indeed, restriction maps are useful metric sequence representations that can be readily tested empirically by using cocktails of commercially available endonucleases. Differences between a theoretical restriction map of a sequence and the size distribution of bands on an agarose gel, when cutting the latter with specific endonucleases, can be used to access the genetic integrity of a mapped sequence. Thus restriction maps provide an easy and straightforward method to screening a segment of DNA for genetic instability, mutations, and recombination.

Following this approach, it was decided that the first step in trouble-shooting pQR493 would be to look at its restriction map. Previous experiments had already indicated that the empirical restriction map did not coincide with the expected

results, but these experiments were base on a very limited number of restriction sites (Figure 17) that did not allow for a detailed assessment of the vector's integrity. Therefore, the initial task would be to complement the predicted restriction map of pQR493 by regressive analysis, i.e. by mapping out the carry-over of restrictions sites from the precursors of pQR493.

However, producing a clear picture of the restriction map was not a straightforward task. Due to the fact that most of the precursor for pQR493 were constructed in the 80s and 90s, and in that period a culture of extensive documentation of genetic constructs was not yet established, we have very limited information about any sequences of these constructs.

Thus, it was only possible to re-create *in-silico* a small portion of the restriction map by analyzing the literature of previous work done on the precursor vectors, thereby rationally drawing up a history of the restriction sites that were maintained and discarded throughout all the engineering steps done to reach pQR493.

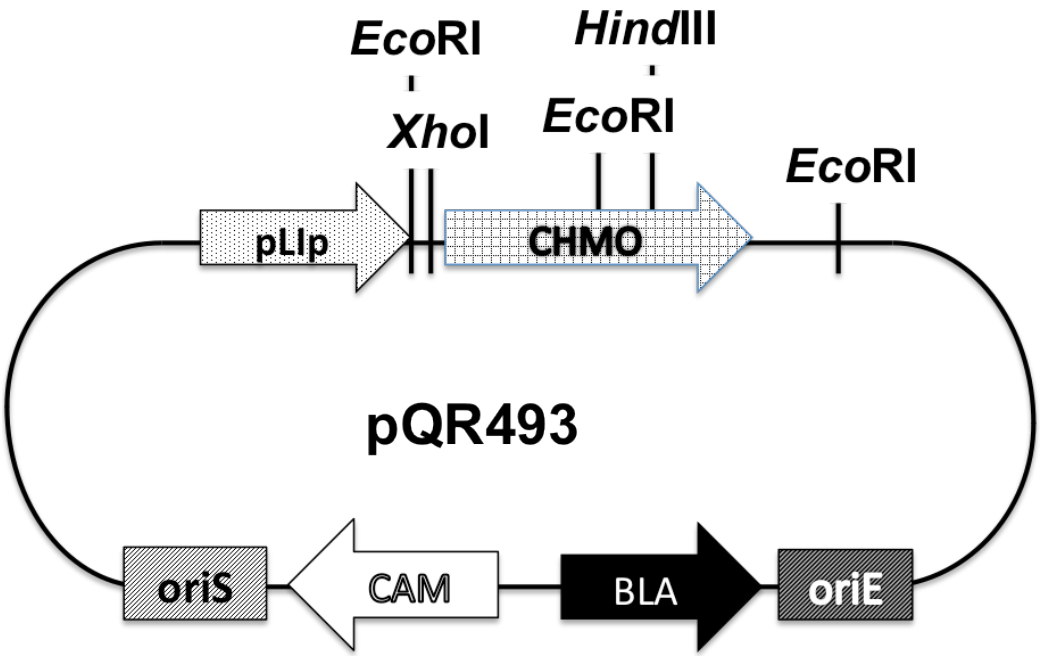


Figure 17- Restriction map of the vector pQR493 as it was available before the *in-silico* analysis of its restriction map.

The *in-silico* analysis of pQR493 focused primarily on the regions around the lipase promoter and CHMO gene, as these seemed to be the most relevant when accessing the expression capability of the vector. If these regions carried mutations or sequence insertions/re-arrangements, the vector's ability to express the biocatalyst could be compromised, and therefore it was imperative to find out if any detrimental events had occurred to the genetic material in the vicinity of CHMO.

The pQR493 vector is a product of four generations of genetic engineering on a couple of basic vector designs, not all of which was done by the same group or in the same study (Figure 18). The starting point (or first generation) was the vector pLipSP1, which in turn is a pC194 derivative in which the lipase promoter and gene from *S. hyicus* were cloned via a *Pst*I linker region. The main use of pLipSP1 was as tool for the study of the lipase promoter and gene in Gram-positive bacteria. Introducing extra restriction sites in the N- and C-terminus of the lipase gene allowed for the identification of a signal-peptide sequence at the start of the lipase gene that promoted secretion and surface display of the lipase protein or any other fusion protein in-frame with the signal pro-sequence of the lipase gene (Demleitner & Götz 1994). This derivative, pLipSP17, representing the second generation of vector constructs, was used subsequently as the basis of several different protein display systems in *S.carnosus*. The restriction sites were introduced into the lipase gene sequence through site directed mutagenesis, and they were a *Bsm*I site at the beginning of the lipase signal sequence, a *Bcl*I site at the end of the signal sequence, and a *Bgl*II at the end of the lipase pro-peptide sequence.

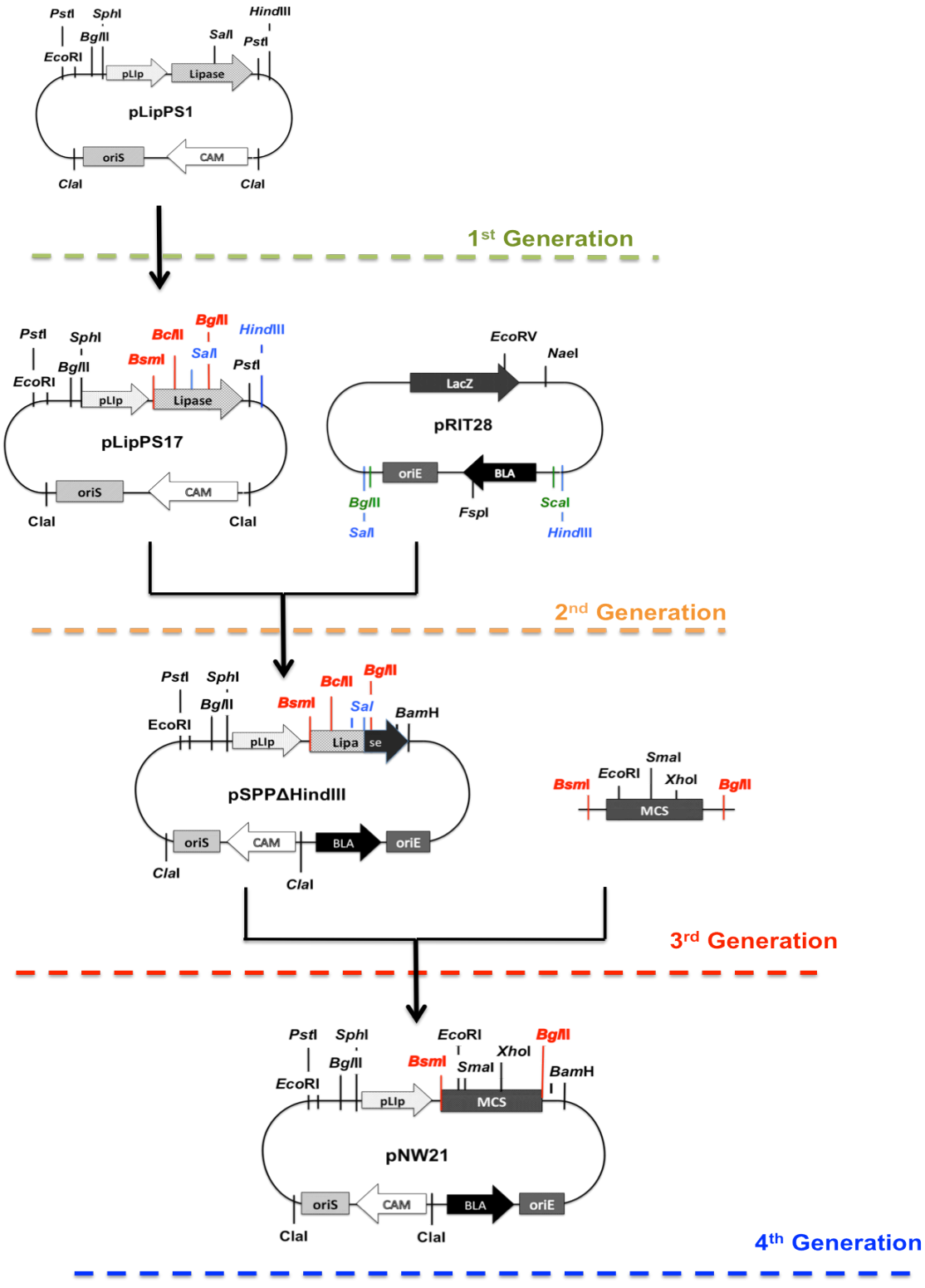


Figure 18- Schematic the genetic history of the vector pNW21. Some intermediate steps are not represented in this diagram. These steps were not considered significant for the resulting restriction map of pNW21, and were therefore disregarded.

The next step in the production of pQR493 was joining the vector pLipSP17 with the replication and antibiotic selection part of a Gram-negative vector, pRIT28, via a *Sal*I-*Hin*dIII restriction and insertion, thus creating a third generation construct that was able to replicate in both Gram-positive and Gram-negative organisms. This fusion vector, pSPP, allowed for a greater flexibility in cloning between the two types of bacteria, and streamlined the genetic manipulation of subsequent vector derivatives by taking advantage of the more simple-to-use *E.coli* cloning system. It is important to note that the insertion of the pRIT28 fragment led to the loss of the C-terminus of the lipase gene, which included the *Bgl*II site previously introduced into pLipSP17. However, this region of the gene was subsequently reconstructed via PCR and re-introduced into pSPP through a SalI-BamHI restriction and insertion. pSPP was then modified to pSPPΔHindIII via a deletion of an HindIII site upstream from the gene, and this derivative was then used in several subsequent shuttle vector constructs.

The last genetic manipulation step that has relevance to the construction of pQR493 was the introduction of a synthetically made multi-cloning site (MCS) into pSPPΔHindIII via a BsmI-BglII restriction and insertion. The resulting vector, pNW21 does not suffer any further significant changes, and therefore most of the restriction map of pQR493 is the same, expect the restrictions sites introduced by the CHMO gene, which will be discussed later. The insertion of the MCS added several unique sites to the restriction map of the vector, most important of which are the *Eco*RI, *Sma*I, and *Xho*I sites.

During the *in-silico* analysis of pQR493's restriction map, several sites were identified that could not be mapped reliably or that conflicted with the reasoning behind the construction of the vectors. For instance, there are two sites in pRIT28, *Bgl*II and *Sca*I, which could not be specifically allocated around the region that is subsequently taken from the vector to ligate with pLipPS17. Since there is no reference to them in the description of pSPPΔHindIII, it was impossible for us to know if these sites were carried over into any of the subsequent generations.

In another instance, a *Bgl*II site upstream of lipase promoter (pLip) that was maintained throughout every generation conflicts with the last step in the pNW21 creation. In this step, the multi-cloning site is insertion into pSPPΔHindIII via a *Bsm*I-*Bgl*II restriction (Williams et al. 2002), but there is no reference to the fact

that *Bgl*II is not a unique site. Cutting the vector in the way the study described would have excised out the region the authors were targeting as well as the pLip region upstream, generating a linear fragment with *Bgl*II over-hangs. The fact that there is no reference to this problem make us assume that either the authors used a partial restriction to clone the MCS, or that the *Bgl*II site upstream has been lost around the third generation of vector precursors.

The insertion of the CHMO gene into pNW21 also introduced a legacy of restriction sites that were not intrinsic to the gene. This was because CHMO was not cloned as a clean PCR-product, but instead was cloned from a precursor vector that had also undergone a series of modifications. Fortunately, all the steps that led to the final version of CHMO were previously well documented in a study that describes the creation of the *E.coli* TOP10 expression vector pQR239 (O'Sullivan et al. 2001).

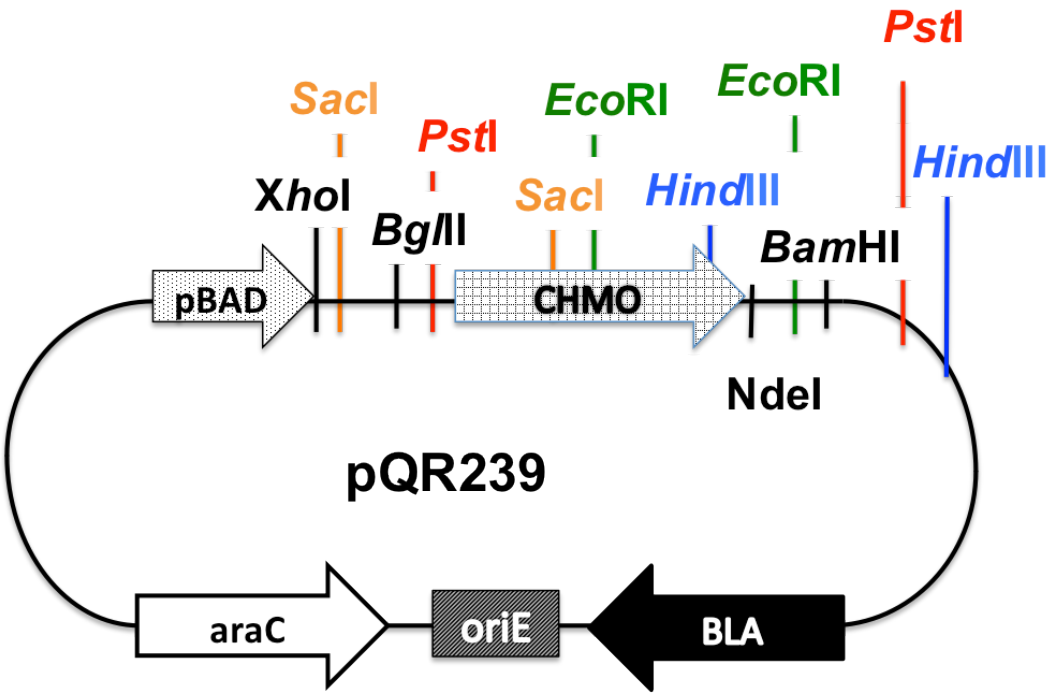


Figure 19- Restriction map of the vector pQR239 taken from O'Sullivan et al. (2001). Black restriction sites indicate unique sites around the CHMO gene, while colored restriction sites indicate the presence of more than one of these sites around or within the gene.

The restriction map of this vector (Figure 19) shows the complexity of restriction sites around the CHMO gene that underpins a major problem of the cloning strategy described in the study. Indeed, the legacy of restriction sites that accumulates after

each ligation step reduce the potential of pQR239 for further genetic engineering by decreasing the population of unique sites that could be used without jeopardising the overall structure of the vector. In the end, for the creation of pQR493 the CHMO gene could only be lifted from its parent vector using the endonucleases XhoI and BamHI, a limitation that partially motivated the insertion of MCS into pSPPΔHindIII in order to accommodate a more flexible region of unique sites.

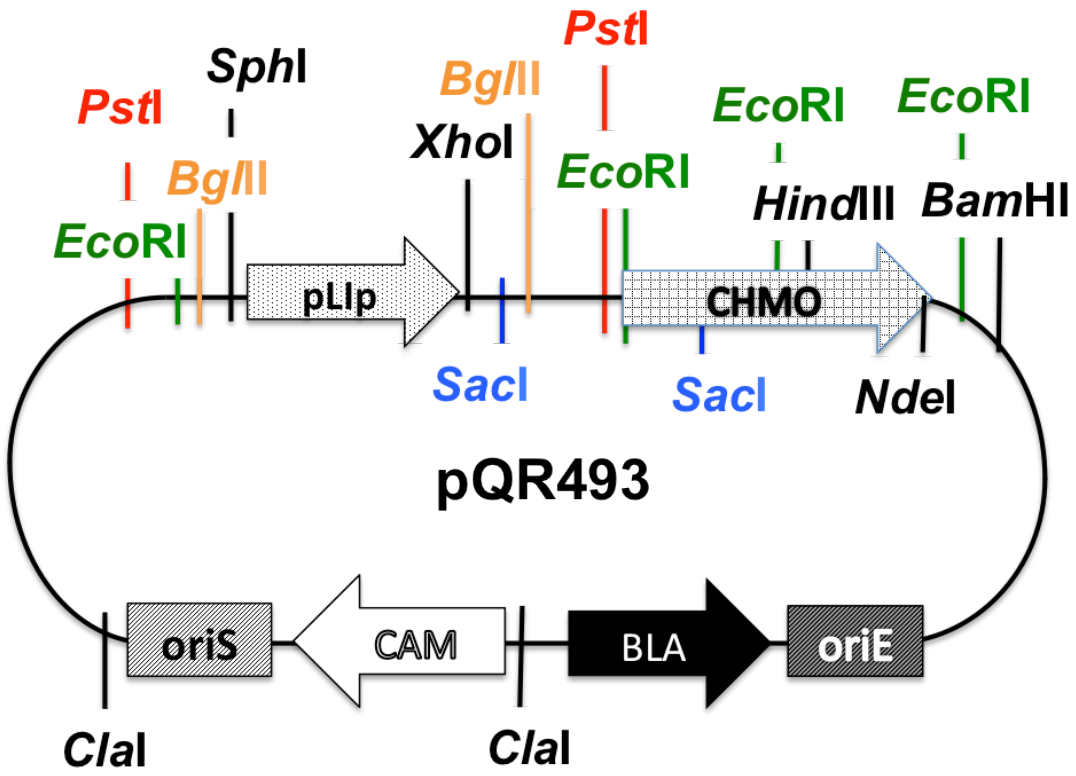


Figure 20- Restriction map of the vector pQR493 after the *in-silico* analysis. The color-coding indicates the uniqueness of restriction sites: black corresponding to sites that are unique within the vector, while the colored sites represent redundant or repeated sites.

With the information from the vector pQR239 and the *in-silico* re-creation of the area around the promoter and MCS of pNW21, a more detailed version of the restriction map of pQR493 was created (Figure 21). The qualitative nature of the representation of the distances between restriction sites is mainly due to the fact that there is very little information about the position of these sites on any of the pNW21 precursors, and therefore it was impossible to confirm the metric distances between different sites.

While the analysis did not provide numerical data for the positions of the restriction sites, it was still crucial for a better understanding of the overall layout of the vector's restriction map and how the different restriction sites were organized upstream and downstream from the promoter and CHMO gene. This information was sufficient to allow for accurate predictions of the experimental outcomes from different restriction digests of pQR493. The added knowledge of the vector also allowed for a greater flexibility in the repertoire of endonucleases that could be used empirically to access the genetic integrity of pQR493.

3.3- Restriction Digests of pQR493

With a comprehensive *in-silico* restriction map of pQR493 constructed, the next step was to plan and perform several empirical restriction digests that would either corroborate or confront the predictions inferred from the map. Consequently, a list of endonucleases was chosen to digest the vector, either individuality or as a cocktail of two endonucleases. The vector was extracted from *E. coli* TOP10 cultures that had been grown overnight in NB2 medium and subsequently lysed using a DNA extraction kit from QIAGEN. The plasmid samples were subsequently restricted by mixing a small volume of plasmid DNA with the desired mix of endonucleases and corresponding reaction buffer. The protocols for plasmid extraction and DNA restriction are described in detail in the materials and methods chapter (see chapter 2, sections 2.2.4 and 2.2.7).

Table 1 shows the list of endonucleases used in the restriction digest experiments, as well as the predictions of how many sites these enzymes would cut in pQR493 when used individually and in combination with a second endonuclease. It's worth noting that not all the combinations shown in the table were performed. Restriction sites like *Bgl*II sites were not used in the empirical analysis because the uncertainty regarding the position of these sites, as it has been explained earlier.

	EcoRI	PstI	XhoI	HindIII	SacI	NdeI
EcoRI	4	6	5	5	6	5
PstI	6	2	3	3	4	3
XhoI	5	3	1	2	3	2
HindIII	5	3	2	1	3	2
SacI	6	4	3	3	2	3
NdeI	5	3	2	2	3	1

Table 1- List of endonucleases used and number of cuts corresponding to the enzymes on their own (white diagonal strip) or in combination with a second endonuclease.

The first round of experimental restrictions focused on the restriction profiles of *Eco*RI and *Pst*I, either used individually or in conjunction with *Hind*III, *Xho*I and each other. All of the endonucleases mentioned were predicted to span most of the region upstream and downstream of the CHMO gene, as well as the middle segment from the latter, and therefore seemed to be best choice to access if any large scale mutation or re-arrangement had happened on that region. Figure shows the results from these digests displayed in an agarose gel, where the vertically arranged bands correspond to different sized linearlized or super-coiled fragments of DNA. Bands located outside of the detectable range window, on the top part of the gel, correspond to undigested DNA in different conformation, while most digested fragments are located within the detectable range. The presence of undigested DNA in lanes is an indication that the plasmid sample was only partially digested. In the case of first experiment run, the heavy presence of undigested DNA was a clear indication that the restriction reactions were sub-optimal.

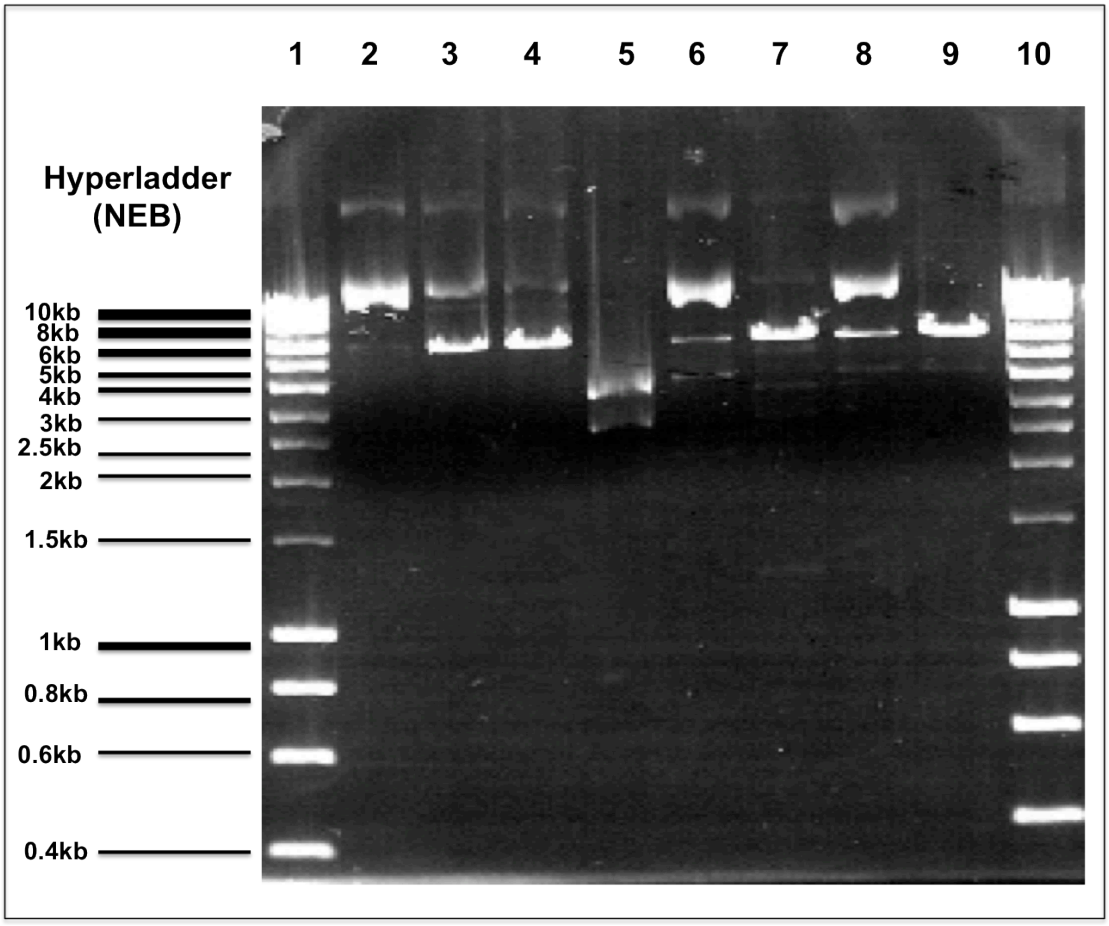


Figure 21- Agarose gel (1% w/v) showing the restriction map of pQR493 using a list of endonucleases. Lane 1 and 10: hyperladder from NEB; lane 2: undigested pQR493; lane 3: *Eco*RI restriction; lane 4: *Eco*RI-*Xho*I restriction; lane 5: *Eco*RI-*Hind*III restriction; lane 6: *Pst*I restriction; lane 7: *Pst*I-*Eco*RI restriction; lane 8*: Pst*I-*Xho*I restriction; lane 9: *Pst*I-*Hind*III restriction.

Despite low performance of the restrcition reaction, the results for the first run could still be analyzed and compared with the *in-silico* restriction map. Some of the results were in accordance with the *in-silico* predictions. For instance, the endonuclease *Pst*I was predicted to cut pQR493 into two not equally sized linearized fragments. This was confirmed by the results on the gel, in which samples digested with *Pst*I digest produced 2 bands of distinct sizes.

On the other hand, the *Eco*RI digests directly contradicted the predictions from the restriction map, as it was expected that this endonuclease would digest pQR493 in at least 4 sites, but instead the gel showed a singular band corresponding to linearized vector. Similarly, a digest combining both *Pst*I and *Eco*RI resulted in linearized vector as well as 3 smaller bands, which suggested

the presence of an unique *Eco*RI site. A double digest of *Eco*RI and *Hind*III seemed to corroborate these results by producing two bands of similar sizes, while also suggesting that the sites for these two enzymes were on almost opposite sides. Most of the *Pst*I digests were very hard to analyse because of the low reactions yields, as evident from the strong bands belonging to linearized vector. Figure 22 shows the revised graphical layout of some of the restriction sites within pQR493, taking into account the results from the first restriction digest experiment.

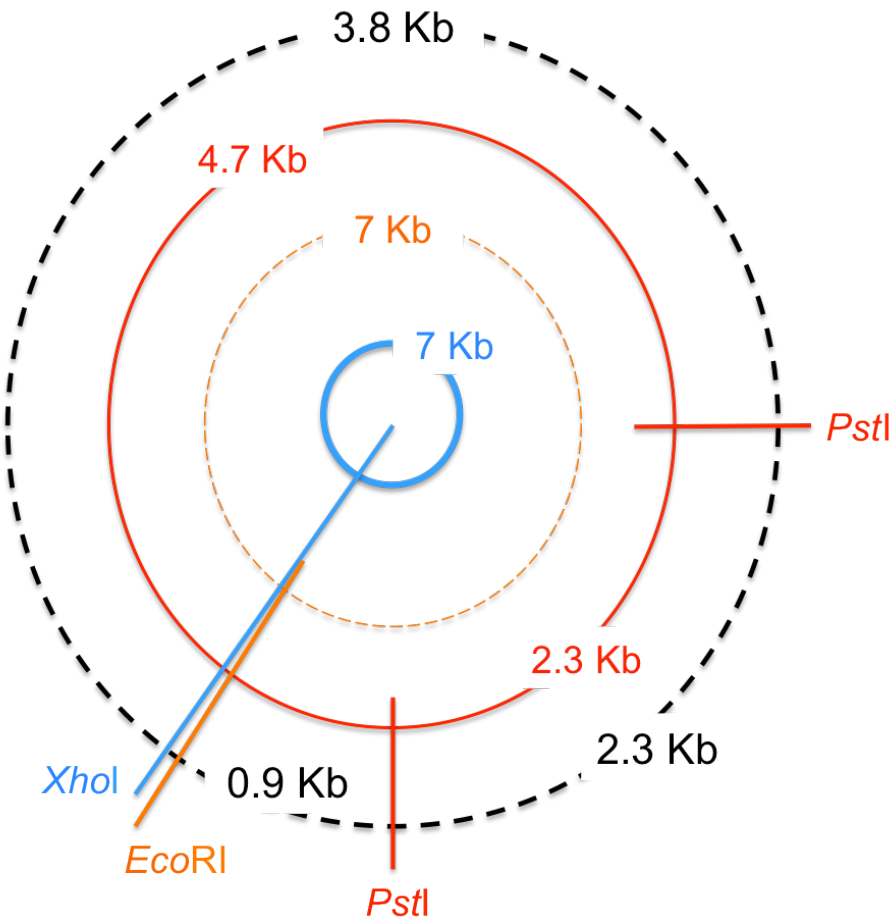


Figure 22- Graphic representation of the sizes and putative positions of the restriction sites, based on the bands from the restriction map experiment. The different colored rings and lines represent the number of cuts produced by the different endonucleases on their own, *Xho*I (blue), *Eco*RI (orange), and *Pst*I (red). The sizes of the bands produced by these cuts are also coded with the same colors respectively. The black ring and sizes refers to the distance relationship between the different sites, and how these could be organized spatially in the vector.

Together, these results suggest that the DNA sequence around and within the CHMO gene suffered from large-scale re-arrangements, deletions, or damaging modifications. In particular, the absence of more *Eco*RI site was rather worrying, as most of the *Eco*RI sites in the restriction map are located within the CHMO gene. However, the agreement between the *Pst*I and *Hind*III restriction digests and the predictions made from the *in-silico* restriction map concerning the latter suggested that at least part of the CHMO gene sequence was still intact, since both the unique *Hind*III site and a second *Pst*I site were located within the gene sequence.

There was also the question of the size of the vector obtained from the restriction digests, which was different from that calculated *in-silico*. Although sizes were not mentioned during the *in-silico* analysis, it was clear that pQR493 should be 8 kb (kilobases) in total, with 6kb being from the pNW21 construct and another 2Kb coming from the CHMO gene and adjacent upstream and downstream overhangs. The experimental results seemed to contradict this prediction, since the cases where the vector was linearized resulted in a single band at about 7 kb, 1kb shorter than the predicted size. However, due to the inefficiency of some of the reactions and the loss of resolution in the higher size ranges of the agarose gel, it was not possible to conclude with certainty that the vector had lost some of its sequence before performing a more detailed and definitive empirical analysis.

A second round of restriction digests was therefore designed, not only to confirm the results from the first restriction experiment, but also as an attempt to acquire a more detailed picture of the sequence around the CHMO and the nature of the damage that might have occurred within the latter. In order to achieve this, the repertoire of endonucleases used in the restriction digest was expanded to include *Sac*I and *Nde*I, as well as the other enzymes used in the previous run. From *in-silco* analysis of the restriction map of pQR493 it was concluded that the sites for both endonucleases had been carried over with the CHMO gene when the latter was ligated to pNW21. Therefore, testing for the presence or absence of these sites in the extracted vector constructs would give a clear indication of the genetic integrity of the CHMO insert.

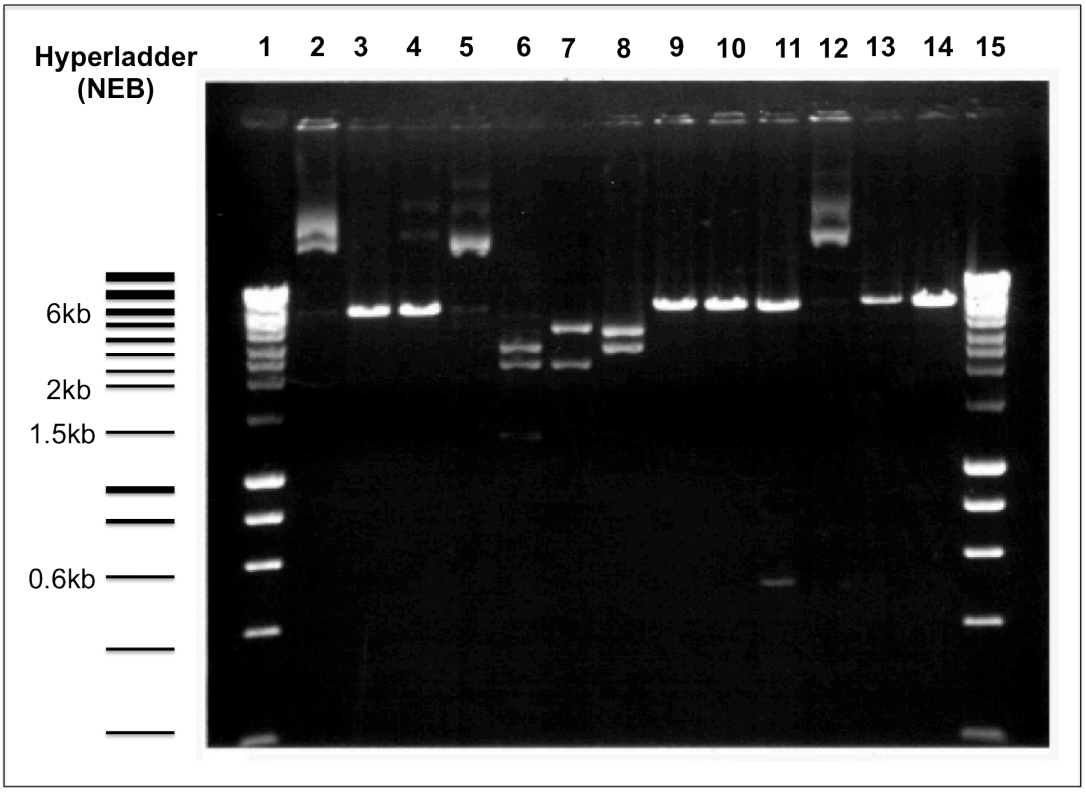


Figure 23- Agarose gel (0.7% w/v) showing the restriction map of pQR493 using a second set of endonucleases. Lane 1 and 15: hyperladder from NEB; lane 2: undigested pQR493; lane 3: *Eco*RI restriction; lane 4: *Eco*RI-*Xho*I restriction; lane 5: *Eco*RI-*Hind*III restriction; lane 6: *Pst*I-*Xho*I restriction; lane 7: *Pst*I-*Hind*III restriction; lane 8: *Xho*I-*Hind*III restriction; lane 9: *Nde*I restriction; lane 10: *Nde*I-*Eco*RI restriction; lane 11: *Nde*I-*Xho*I restriction; lane 12: *Sac*I restriction; lane 13: *Sac*I-*Nde*I restriction; lane 14: *Sac*I-*Hind*III restriction.

Figure 23 shows the results of the second set of restriction digests. The results from the *Eco*RI digests corroborated the data taken from the first run of restrictions, by producing what could be interpreted as a linearized vector when using the enzyme by itself or in the presence of *Xho*I. These results seemed to suggest that the *Eco*RI and *Xho*I sites were located close together in the backbone of the vector.

On the other hand, the fact that a *Xho*I/*Hind*III restriction produced two bands of similar sizes corroborates the previous data showing that *Xho*I and *Hind*III sites stood at almost opposite sides of the vector, judging by the close size proximity between the two bands. The calculated distance between the two sites was far greater than the expected distance if the CHMO gene was indeed inserted into the vector. Instead, the positioning of these two sites within pQR493 was more

analogous to the distance between the XhoI site of pNW21 and the HindII site of pSPPΔHindIII prior to its deletion.

The *Pst*I digests resulted in good reaction efficiencies in the second run of restriction and corroborated the results from the first run, while also adding some important information. On one hand, the similar restriction profile produced by a *Xho*I/*Pst*I restriction when compared with a *Eco*RI/*Pst*I digest performed in the previous run re-enforced the conclusion that the *Eco*RI and *Xho*I site were located in close proximity. On the other hand, the *Hind*III/*Pst*I digest produced two distinctly sized bands, instead of the three bands that had been predicted from the two *Pst*I sites and unique *Hind*III site located around the CHMO gene. This result suggested two possible outcomes: either the *Hind*III was in close proximity to one of the *Pst*I sites, and therefore the distance between these sites was below the range of detection of the agarose gel; or the *Hind*III was non-existent. However, the results from the *Xho*I/*Hind*III restriction contradicted the latter possibility, by showing that *Hind*III contained a unique restriction site in pQR493.

The *Sac*I digests showed the absence of any *Sac*I sites in the vector, thus suggesting that the N-terminus of the CHMO was absent altogether. On the other hand, the *Nde*I digests still produced linearized vector when the endonuclease was used individually, suggesting the C-terminus of the gene was intact. Alternately, combining this enzyme with either *Eco*RI or *Xho*I suggested that there was a distance of 500 bp between *Eco*RI and *Xho*I, with *Nde*I being in close proximity to *Eco*RI. This distance relationship had not been observed in any of restrictions with the other enzyme combination, and it did not fit with the results from the *Xho*I/*Eco*RI digests.

Finally, the size of the linearized vector obtained in the second run of experiments was still smaller than the predicted size of 8Kb for pQR493. Figure 24 shows a graphical representation of the revised layout of restriction sites in pQR493 after the second restriction experiment.

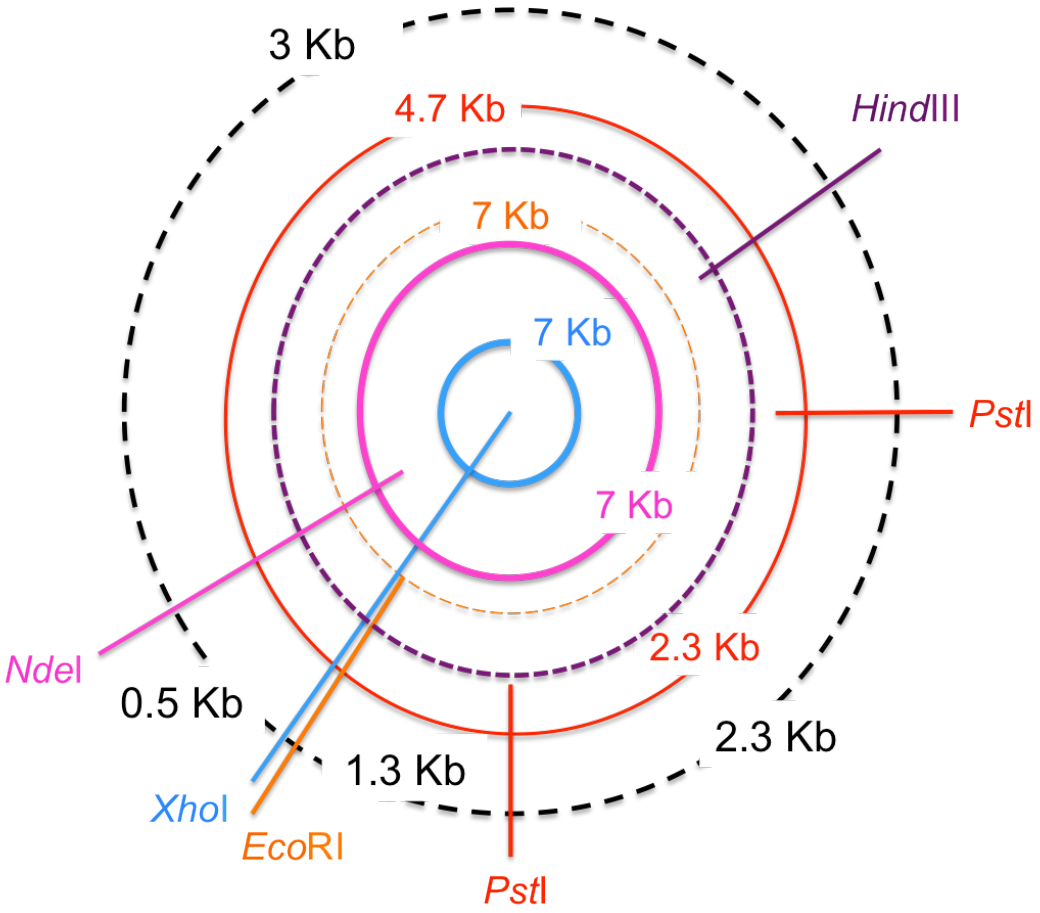


Figure 24- Graphic representation of the position of different restriction sites based on the restriction bands from the second set of digests. The colored rings and lines represent the number of cuts that different endonucleases generate in pQR493. The sizes of the bands corresponding to the cuts are also represented by the same colors. The band ring and numbers refer to the distance relationship between the different restriction sites, and how these could be spatially organized in pQR493.

Together, the results from both experimental sets seemed to suggest one of two possible scenarios. In a worst-case scenario, most of the N-terminus of the CHMO gene, where the *Sac*I sites should be located, had been deleted. This hypothesis was hard to fit with the overall restriction layout obtained from the results. In particular the *Hind*III site, which would be located in the C-terminus of CHMO, was very distant from the *Xho*I site, which in turn was located downstream of the Lip promoter as part of the remaining MCS from pNW21. If it were the case that only a truncated version of the CHMO gene sequence containing *Hind*III and *Nde*I managed to ligate into the MCS of pNW21, then one would expect the *Xho*I and *Hind*III sites to be in close proximity. The more likely scenario would be that

the vector pQR493 was in fact a version of pNW21 with an active *Hind*III site. This hypothesis seemed to fit better with the results, except for the *Nde*I site, which had not been predicted as intrinsic to pNW21 in the *in-silico* analysis. The close proximity of the *Xho*I and *Eco*RI sites would be explained by the MCS of pNW21. In addition, since the position of the second *Eco*RI site that lies upstream of the lipase promoter is not know, the close proximity of between this site and the *Eco*RI site located in the MCS region in pNW21 could explain why samples cut with this endonuclease failed to produced two discernible bands. However, the positive activity shown with *Hind*III was not expected in the case of pNW21, and suggested that the *Hind*III site from the precursor vector pSPP had not been deleted, contrary to previous reports of the vector.

Regardless of which scenario corresponding to the truth, it was clear that the integrity of the CHMO gene was compromised and consequently the available vector construct could not be use to express the biocatalyst in *S. carnosus*. Therefore the project needed to re-trace a step to create a functional expression system.

3.4- Ligation Attempts with pNW21 and CHMO

Once it was clear that the vector that had been provided from previous work on the project could not be used to express the biocatalyst CHMO in *S. carnosus*, the focus of the project shifted from the characterization of whole-cell biocatalysis using *S.carnosus* to the production of an expression system that would allow the intracellular expression of CHMO in the Gram-positive bacteria.

The simplest and fastest way to achieve a working genetic system would be to repeat the cloning strategy that was previously used in the creation of pQR493. Accordingly, stocks of *E.coli* TOP10 containing the vector pNW21 and CHMO expression vector pQR239 were re-grown to create banks of DNA for use in the subsequent genetic manipulations.

As described above, the strategy used for the insertion of the CHMO sequence into pNW21 involved two steps: the extraction the CHMO gene from pQR329 through a *Xho*I/*BamH*I restrictions, and subsequent ligation with the *Xho*I/*Bgl*II restricted pNW21. Four of the six nucleic acids from the overhangs of *BamH*I and *Bgl*II are complementary to each other, and therefore can be ligated together, resulting in a sequence that does not contain a recognition site for either endonuclease. The choice of different endonucleases to ligate the C-terminus of CHMO to pNW21 seems peculiar in hindsight, especially after the *in-silico* analysis of pNW21 identified the presence of a unique *BamH*I site downstream from the MCS. Therefore, in this project two different ligation protocols were attempted, in which pNW21 was alternately restricted with *Xho*I/*Bgl*II and *Xho*I/*Bam*HI mixes prior to ligation with the *Xho*I/*Bam*HI restricted CHMO gene.

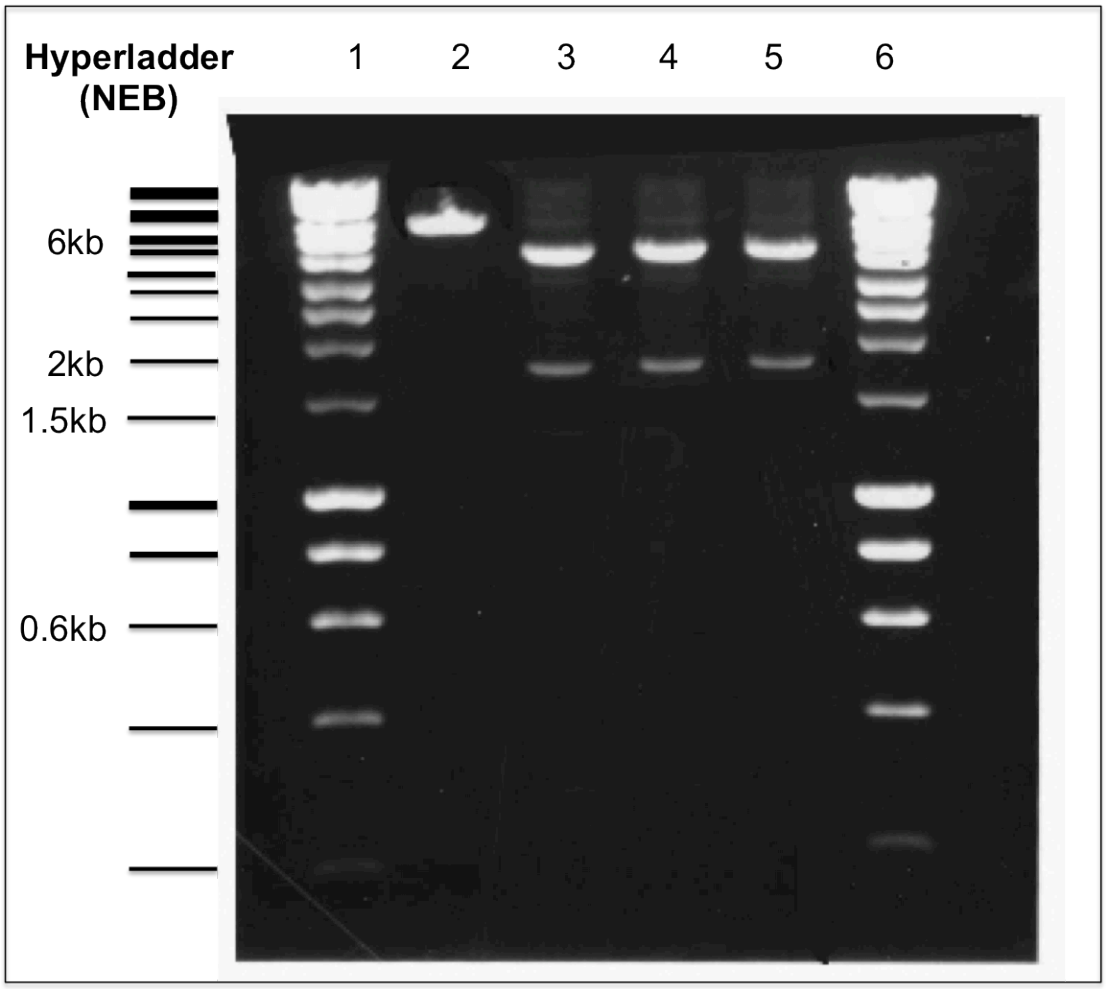


Figure 25- Agarose gel (0.7% w/v) showing a restriction digest of the pNW21 and pQR239 with XhoI-BglII and XhoI-BamHI respectively. Lanes 1 and 6: DNA hyperladder (NEB); lane 2: pNW21 digest; lane 3 to 5: pQR239 digests. Cutting pNW21 generated a single band corresponding to the size of the pNW21 vector, while cutting pQR239 generated two bands, the lower of which corresponds to the CHMO gene (1.6 kb).

Figure 25 shows an example of the digested plasmid DNA sequences from pNW21 and pQR239 displayed on an agarose gel. Restriction samples were loaded onto an agarose gel in order to screen the samples for high populations of un-cut vector or re-ligated vector, which would hinder the ligation process by increasing the number of false positives after transformation (i.e. the number of colonies that contained the pNW21 vector without the CHMO insert). In the case of the pNW21 digestion, it was difficult to access if it had been cut properly with both restriction endonucleases, since the restriction sites for these enzymes were very close together. Therefore, we could only indirectly screen for efficient pNW21

digestion by running it in parallel with the pQR239 digestion, which used the same endonucleases, and could be easily accessed in a gel because of the two bands it produced. The smaller of the two bands corresponded to the CHMO insert (1.6 kb in size), which had to be excised and extracted from the agarose gel prior to ligation.

The extraction of restricted fragments from gels was preferred to using the digested fragments straight after restriction partly because gel extraction was required to separate the CHMO fragment from the parent vector pQR239, and partly because after endonuclease inactivation via heat treatment there is the risk of samples containing residual enzyme activity, which could re-digest the ligated fragments. Thus, both the pNW21 vector and CHMO gene were purified from gels and subsequently ligated.

There are no universal rules for a successful ligation protocols. Some methods can vary depending on the enzyme commercial provider in terms of incubation times, volumes, and concentrations, while other vary depending on the quality and type of DNA used. In the case of the ligation strategy used for the ligation of CHMO to pNW21, different commercial variants of ligase were used in an array of conditions with varying vector and insert concentrations and volumes, as well as varying ligase concentrations. A detailed description of the different ligation protocols used in this project can be found in the materials and methods chapter (chapter 2, section 2.2.3).

Figure 26 shows an example of a ligation experiment with T4 ligase from NEB (New England Biolabs). This agarose gel demonstrates that after ligation the band profile of a sample containing a mix of CHMO and pNW21 fragments changes, and this change indicated that there was some degree of ligation between fragment and vector, as well as re-ligation of vector.

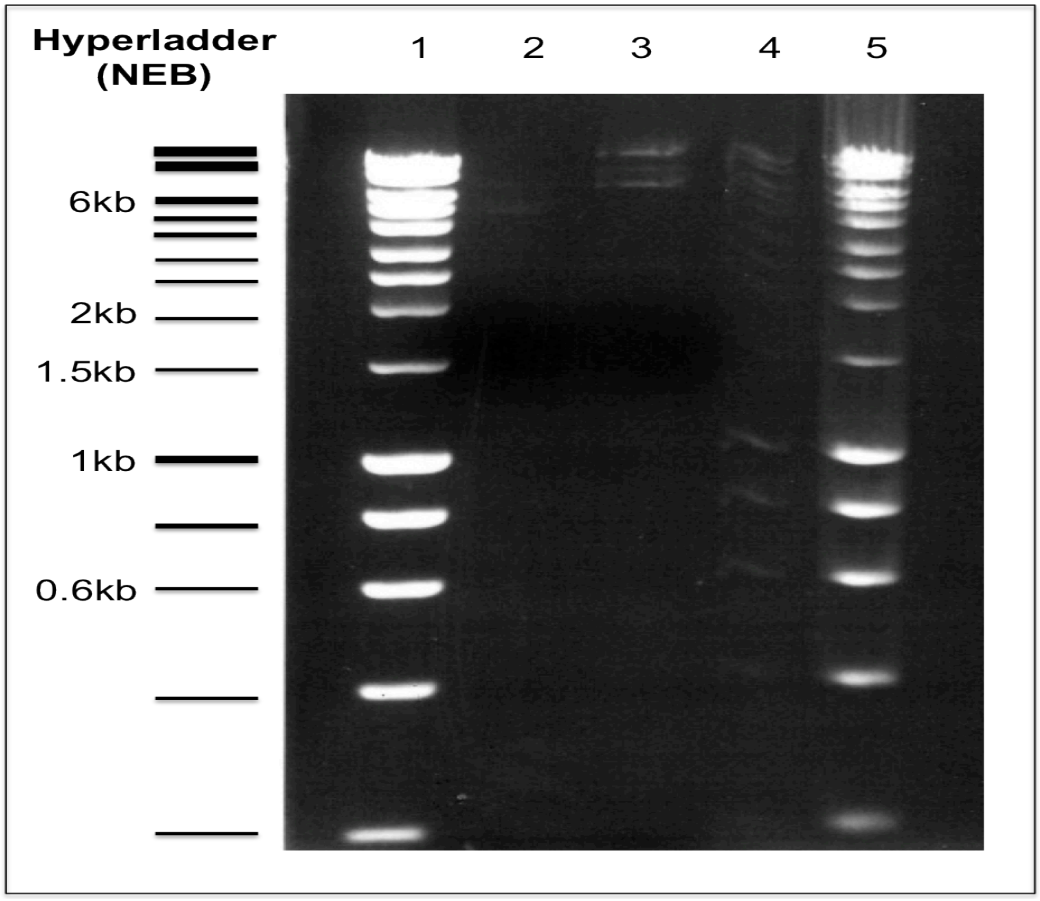


Figure 26- Agarose gel (0.7% w/v) showing the ligation o the CHMO gene to pNW21. Lanes 1, 4, and 5: DNA hyperladder (NEB); lane 2: un-ligated sample; lane 3: ligated sample. The un-ligated sample contained both the linearized pNW21 and the CHMO gene fragment. The difference between the sample before and after ligation indicates that a change occurred after addition of T4 ligase.

The concentrations of DNA fragments used varied greatly between ligation experiments. As a rule a thumb, a molar ration of 1:1 between vector and insert was considered to be optimal for ligation to occur, as there would be the same number of molecules from each fragment in solution. However, this ratio was routinely increased as an attempt to switch the reaction equilibrium in favor of the ligation of CHMO with pNW21, as opposed to re-ligation of the vector.

Ligation was stopped through heat-shock at 65 ^{0}C, and the ligation samples were subsequently transformed into competent cells, which would be selected on antibiotic containing agar plates for positive ligations. In theory, because of the selective pressure introduced by the antibiotics, only colonies containing the pNW21 vector would survive on the antibiotic due to the presence of a resistance

gene in the vector. On the other hand, the restriction of pNW21 prior to ligation with two different endonucleases served a secondary purpose of preventing re-annealing of the vector ends in the absence of the CHMO fragment, and therefore no re-circularized plasmid should be present in surviving colonies. In practice, many of transformation experiments with the ligation mixes resulted in a large population of surviving colonies on the antibiotic selection plates. Additionally, all of the colonies that were subsequently screens were shown to contain the re-ligated pNW21 instead of the desired ligation product. The high rates of vector re-ligation suggested that the restriction of the latter was not efficient. However, since the vector was still appearing as a linearized fragment in the agarose gels, and the same endonuclease mix was efficiently cutting CHMO, it was difficult to assess which of the endonucleases was not cutting properly.

Screening of colonies was done by re-growing them overnight and extracting the plasmid DNA. The plasmid samples were subsequently cut with specific restriction endonucleases to check for a restriction map that corresponded to the ligation product pQR493. For the ligation samples in which pNW21 had been digested with a *Xho*I/*Bam*HI mix, the same endonucleases were used to screen for the insertion of the CHMO fragment. However, all the colonies screened with this endonuclease mix resulted in linearized plasmid, indicating that the insert had not been successfully ligated. For the ligations in which pNW21 was restriction with a *Xho*I/*Bgl*II mix, the same screening strategy could not be used, since the resulting ligation would not have either a *Bgl*II or *Bam*HI site. Instead, a *Xho*I/*Hind*III restriction was performed to screen colonies from these ligations, since the *Hind*III endonuclease would only cut the vector if the CHMO gene was inserted. Figure shows the predicted restriction map from the *Xho*I/*Hind*III restriction of the ligation product pQR493, as well as the results from a screening experiment in which the same endonuclease mix was used. Cutting the plasmid from selected colonies produced a very different picture from the predicted restriction map, which was concerning for a number of reasons. Firstly, all colonies that were screened using this protocol showed the same profile, which indicated that either all the plasmid constructs were ligation products, or more realistically, that the *Xho*I/*Hind*III digestion was unable to select between ligation and plasmid re-circulation of pNW21. The second, and perhaps more concerning reason, was the fact that the

restriction map produced by the cut was very similar to the restriction profile of the "faulty" vector pQR493 previously studied when cut with the same enzymes (Figure 27).

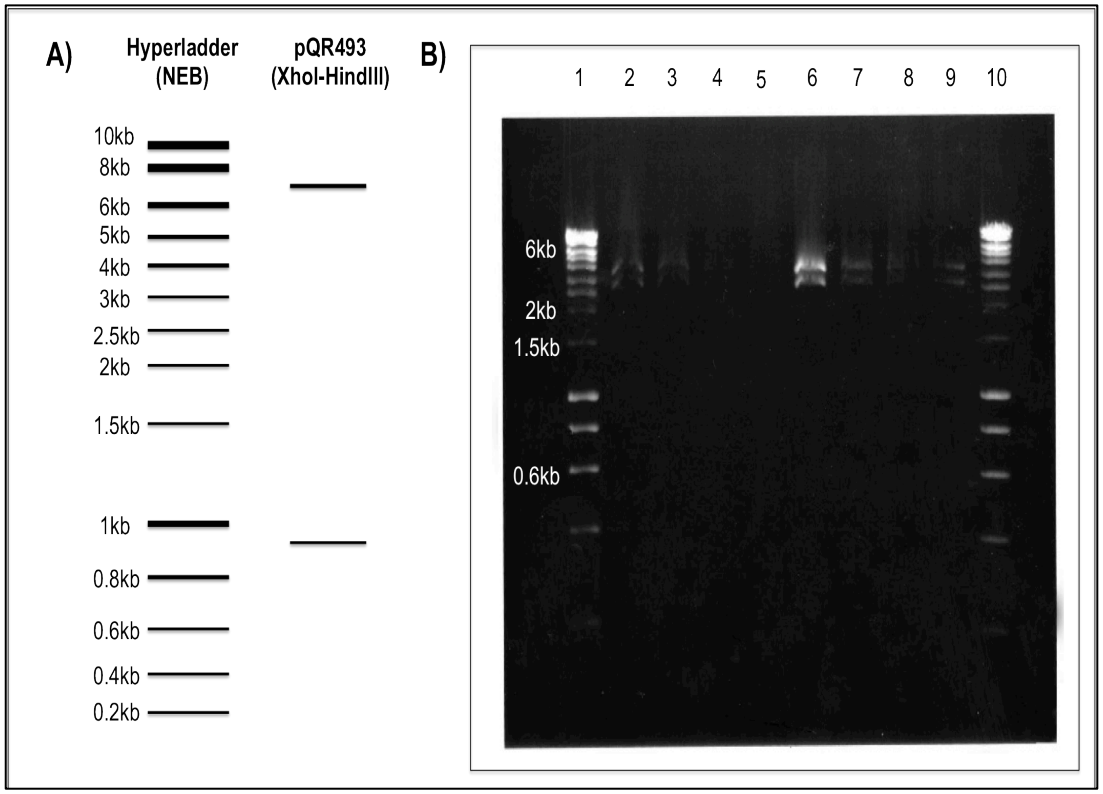


Figure 27- Screening of *E.coli* colonies for ligation between pNW21 and the CHMO gene. A) Predicted restriction map of the ligation product pQR493 when cut with *Xho*I and *Hind*III; B) Empirical *Xho*I/*Hind*III restriction digest of plasmid DNA isolated from colonies that had been selected on ampicillin-containing agar plates. Digested samples were loaded on a 0.7% (w/v) agarose gel. Lane 1 and 10: DNA hyperladder (NEB); lanes 2 to 9: plasmid samples digested with *Xho*I/*Hind*III.

The results from the screening assays of the pNW21/CHMO ligation suggested that *Hind*III site that was detected on the "faulty" pQR493 could have originated from the pNW21 backbone instead of as mutation randomly occurring during the previous ligations of pQR493. To test this hypothesis, pNW21 stocks that had been used in the ligation protocols were restricted with a *Xho*I/*Hind*III reaction mix. The results confirmed that pNW21 produced the same restriction profile as the one generate with pQR493 during the restriction map studies, and subsequently that a HindIII site was located in the backbone of pNW21, when there

should not be one. On the other hand, these results confirmed that the pQR493 construct used in the beginning of this project was in fact the re-ligated pNW21 plasmid.

The presence of an unexpected *Hind*III site and the inability to clone the CHMO into the vector led us to believe that problems first encountered with pQR493 had been a legacy from previous generations of vector precursors, and most likely pNW21 has been a carrier to this legacy. Therefore, before we could carry on with the re-construction of pQR493, it was imperative to assess if pNW21 was still fit for purpose.

3.5- Sequencing of pNW21

It is most likely the erroneous construction of pQR493 stemmed from the presence of the *Hind*III site in pNW21. Since the precursor vector pSPPΔHindIII had been created by deleting the *Hind*III sites in the plasmid backbone, in theory cutting pNW21 with the endonuclease should not have produced linearization of the plasmid. Thus when pQR493 produced a 2-band restriction map after digestion with *Xho*I and *Hind*III, it was assumed that the vector was the product of the ligation between pNW21 and CHMO gene, which contained a unique *Hind*III site within its sequence.

The presence of the *Hind*III site in the backbone of pNW21 posed a serious question: was the restriction site simply omitted from the literature, or was it symptomatic of genetic damage that occurred to the pNW21 stocks. And in the latter case, there was also the question of the extent of the damage . However, after the exhaustive study of pQR493 using a restriction map analysis, repeating this methodology would not add to the understanding of pNW21, and therefore it was decided that a more powerful approach was needed to answer these questions. This approach was to use AB sequencing technology to sequence segment of the vector that were directly related to its functionality.

Figure 28 shows a representation of the strategy used for the sequencing experiments. Primers were designed to anneal with the ampicillin resistance gene

(BLA), the chloramphenicol resistance gene (CAM) and the MCS, and to allow the sequencing of DNA regions within and between these sites. One reason for this choice of primer location was the fact that the sequence from these primers would cover regions that are crucial for the functionality of the plasmid. In addition, the primer design strategy was limited to areas of the vector with known sequences, which could be subsequently compared with the resulted from the sequencing experiments. The sequencing experiments were outsourced to the UCL Wolfson Institute sequencing departments.

Results came in stretches of 700 bp for each primer, which were subsequently screened for overlapping areas and correspondingly joined as longer segments of the pNW21 vector. These longer sequences were subsequently aligned to known areas of the vector in order to identify any mutations. Several software that allow for DNA manipulation and alignment were employed for this this analysis. All the fragments and alignments are shown in the appendix section (Appendix 1).

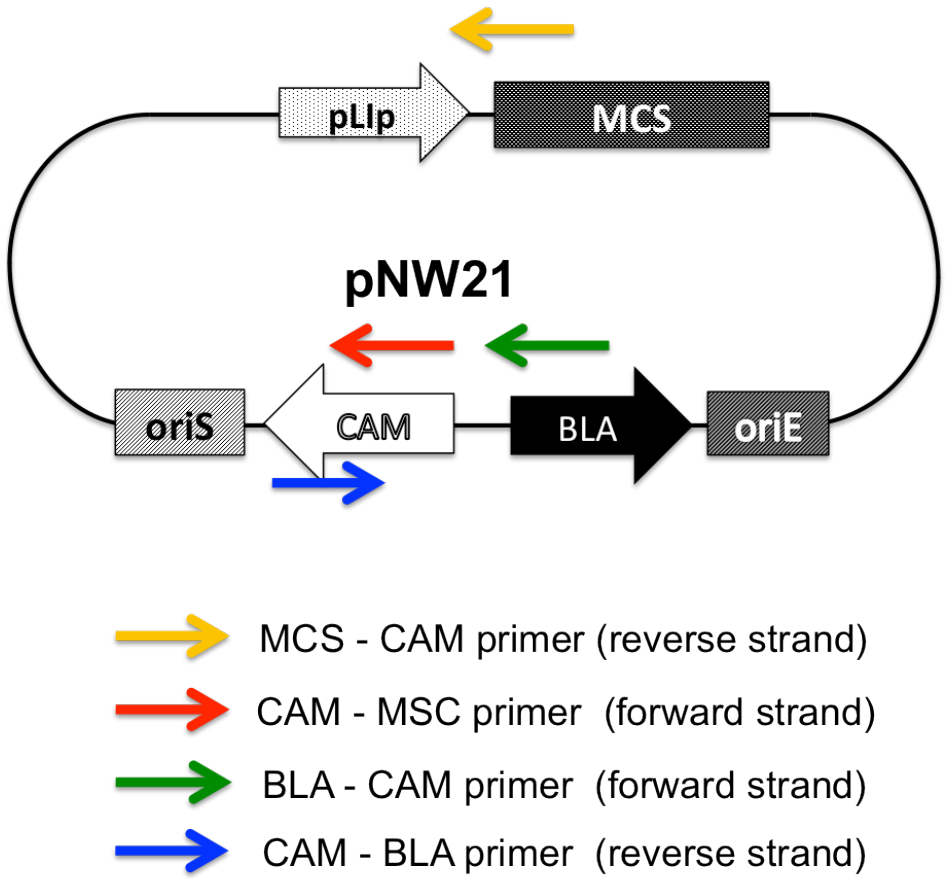


Figure 28- Schematic of the pNW21 vector with the colored arrows representing the primers designed for sequencing purposes. The arrows point towards the orientation of the sequencing from each of the primers. Since sequencing was conducted from 5' to 3', the direction of the CAM-BLA primer (blue) and the MCS-CAM primer (yellow) were designed to allow sequencing of the complement strand of pNW21.

Using this method, a continuous stretch between CAM and the MCS was successfully sequenced, as well as part of the CAM gene. The stretch of DNA sequenced from the BLA gene did not align with any of the other sequences, probably due to the fact that the BLA gene is probably too distant from the other known functional elements. Despite this, the location of the *Hind*III site was pinpointed to this sequenced segment of DNA upstream of the BLA gene. However, because the sequence did not align correctly with the vector pC194, from which the *Hind*III site was originated, it was concluded that the restriction site was not a legacy from any of the pNW21 precursors, but rather introduced by mutation.

On the other hand, most of the sequence between MSC and CAM aligned with the vector pC194, a result that confirmed the ancestry of the latter as a precursor to pNW21. The analysis of the restriction map of this sequence also highlight the location of the *Nde*I site in the stretch that aligned with pC194. This result corroborated the data from the restriction map experiments on pQR493, in which a *Nde*I was detected, despite the absence of the CHMO gene.

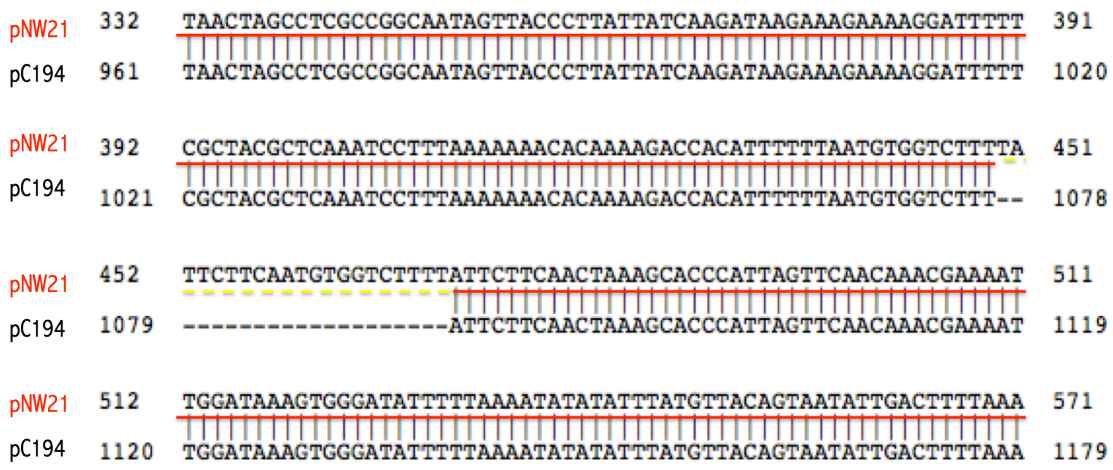


Figure 29- Short segment of the alignment between the sequenced region of pNW21 between the CAM gene and the MCS and pC194. The starting codon of the chloramphenicol resistance gene (CAM) is located at position 1260 in the vector pC194, just upstream of the displayed segment.

The alignment between the sequenced DNA fragment and pC194 also highlighted an unexpected insertion that had occurred just before the CAM gene. As Figure 29 shows, the fragment from pNW21 contained a thymidine (T) –rich sequence in the region upstream of the CAM gene that was not present in the parental vector pC194. The fact that the alignment of sequences upstream and downstream of this segment was 100% accurate validated this sequence insertion as a positive result rather than an error from the sequencing data.

The CAM gene is a crucial functional element of the pNW21 vector due to the fact that it is the only selection factor for the replication of the vector in *S. carnosus*. The loss of function of this gene is fatal to the functionality of pNW21 as an expression system for the Gram-positive bacteria, and therefore it was

imperative to access if the sequence insertion upstream of the gene was detrimental for its expression. For this purpose, colonies of *E.coli* TOP10 carrying pNW21 were grown overnight on ampicillin selective medium and subsequently loaded unto chloramphenicol selective plates with varying concentrations of chloramphenicol (from 5 to 15 µg/ml), which were incubated for 24 hours at 37 ^{0}C. After incubation, the number of surviving colonies was counted as a measure of CAM gene expression. Unfortunately, no colonies survived on the chloramphenicol selective plates after incubation, a result that was corroborated by subsequent growth experiments in which the *E. coli* stocks containing pNW21 failed to grow on NB2 medium with 15 µg/ml of chloramphenicol. From these results it was concluded that the sequence insertion observed upstream of the CAM gene knocked-out the expression of chloramphenicol resistance in *E. coli*, therefore severely impairing the use of pNW21 as an expression system for *S. carnosus*.

3.6- Concluding Remarks

In this chapter we described a legacy of problems that had been carried over from previous work on the staphylococcal expression systems pNW21 and pQR493. Most of the trouble-shooting and mistakes that permeated the start of this project could have been avoided if these expression systems had been previously well characterized. For instance, the miss-interpretation of pNW21 as pQR493 originated from predictions that did not account for the presence of a *Hind*III site in the backbone of pNW21. We were also led to make the same mistake on the basis of an erroneous pQR493 restriction map, which was subsequently rectified through the empirical characterization of the vector.

Since most of the history around the creation of pNW21 is not well documented, the possibility that an extra *Hind*III site had not been accounted for in the restriction maps cannot be disregarded. However, the sequencing results suggest that the genetic integrity of the vector might have been compromised through mutations and insertions of foreign DNA. Such events were damaging

enough to knockout the activity of the CAM gene, which in turn terminated the ability of pNW21 to operate in *S. carnosus* as an expression system.

Therefore, after trouble-shooting pQR493 and pNW21, we were left with the clear understanding that the vectors we were given were not useful as expression systems in *S. carnosus*, and that we would have to either re-trace even more steps and work with a precursor from pNW21, or create an expression system *de-novo*. As it has been thoroughly demonstrated in this chapter, the possibility of working with a precursor from pNW21 would not solve the problem of having to work with unknown or badly characterized genetic material. On the other hand, creating an expression system *de-novo* would allow us to have full understanding and control over the genetic backbone of the system while granting the flexibility to re-arrange and change genetic elements freely without compromising its overall functionality. We decided on the latter option, by re-designing an expression system that would allow expression of biocatalysts in both Gram-negative and Gram-positive bacteria.

Chapter 4: Results- Design, Construction, and Implementation of a Novel Expression System for S. carnosus.

4.1- The Design Strategy behind the Construction of a Novel Expression Vector

There were two reasons driving the decision to re-design an expression vector for *S. carnous.* On one hand, to avoid working with badly characterized genetic systems that would require some level of characterization prior to their use, as the previous chapter demonstrated. On the other hand, a total re-design would allow for a greater flexibility in the choice of the genetic elements making up the vector. Nevertheless, the design would still be limited by functionality constraints. As an interchangeable expression vector between Gram-negative and Gram-positive bacteria, the design needed to cater for replication and selection in both bacterial types. In addition, it would also have to allow for protein expression of the biocatalyst in both *E. coli* and *S. carnosus*.

The idea of creating a shuttle vector for cloning and protein expression in both Gram-negative and Gram-positive bacteria is not a novel concept. For instance, there are several examples in the literature of shuttle vectors created for heterologous protein expression in *S. carnosus* that also allow for cloning in Gram-negative strains. One of these examples is the vector pSPPmABPXM, initially created as an expression system for protein surface display (Samuelson et al.

1995), and more recently adapted to express intracellular proteins under the influence of the constitutive lipase promoter (Williams et al. 2002). While these vectors could be used as functional expression vectors for the purposes of this project, they also had several limitations. One of these limitations is the inability to express heterologous proteins in Gram-negative strains, which means that successful cloning of a biocatalyst to the vectors could only be tested upon transformation into *S. carnosus*. Since the cloning step would be performed in the Gram-negative *E. coli* TOP10, due to the fact that it is a very well characterized cloning strain, it would be useful if the vectors allowed for expression of the cloned product in the latter. Another, perhaps more important limitation is the inherent low degree of flexibility of these vectors relative to the cloning of heterologous proteins. In fact, pSPPmABPXM has a very low number of restriction sites that can be used for protein cloning due to the fact that it is a product of several generations of cloning events with DNA sequences containing homologous restriction maps.

Figure 30 illustrates the modular approach adopted for the re-design of the expression vector, under which the different genetic elements (modules) were reversibly linked together through the use of a synthetic linker. This linker was designed to fulfill several functions: firstly, to allow for the ligation of the different genetic elements together, thereby removing the need for these elements to have homologous restriction sites; secondly, to provide for the cloning of heterologous proteins by the introduction of a MCS with a range of unique restriction sites; finally, to cater for the expression of the cloned proteins through a combination of different removable promoters from both Gram-positive and Gram-negative bacteria. By creating an independent unit for cloning and expression of heterologous proteins, this approach allows for a greater flexibility in the choice of the different genetic elements for replication and selection in Gram-positive and Gram-negative strains. Moreover, the reversible nature of the design ensured that the different components of the vector could be re-arranged or changed depending on the functional needs of the expression system, thereby resulting in a greater adaptability to different bacterial hosts.

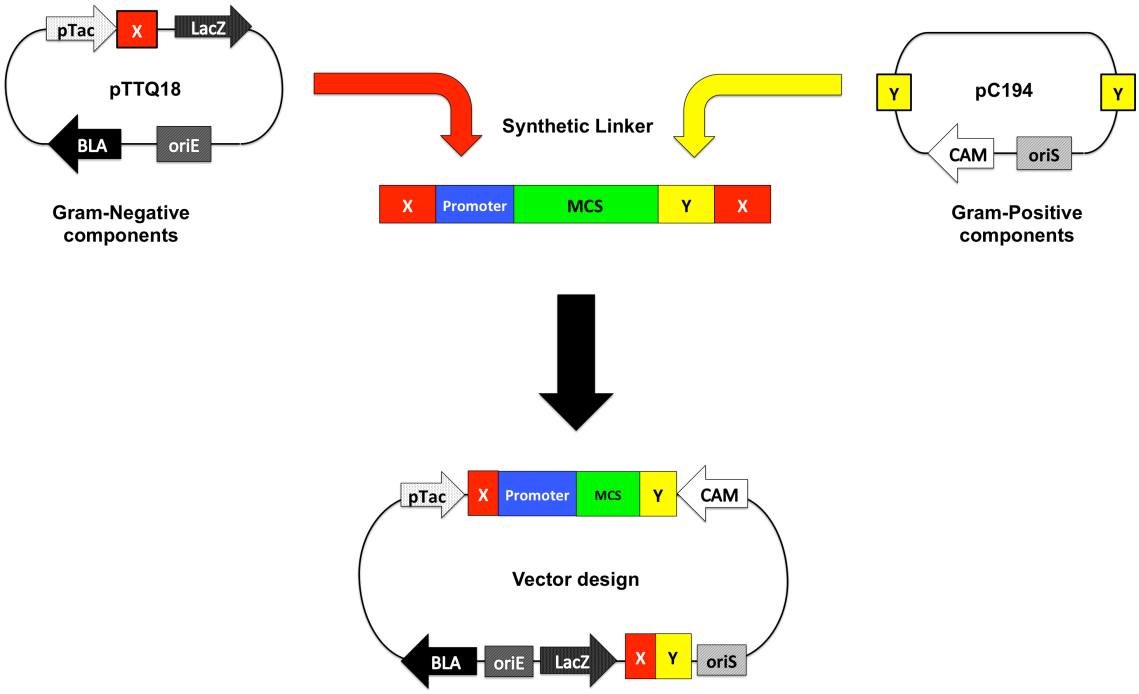


Figure 30- Strategy for the design of a novel shuttle vector for protein design. The synthetic linker contains elements that allow for the sequential assembly of the Gram-negative and Gram-positive components. The ends of the linker (red) are initially connected to the Gram-negative components, after which the Gram-positive components are connected via a restriction site within the synthetic linker.

Similarly to other vector designs previously constructed for *S. carnosus*, it was decided that the staphylococcal plasmid pC194 would work as the source of replication and selection genetic elements for the Gram-positive bacteria. This 2.9 kb plasmid was first isolated from *Staphylococcus aureus* as a small double-strand circular DNA that encoded chloramphenicol-induced-resistance through the expression of the enzyme chloramphenicol acetyl transferase (CAM) (Horinouchi & Weisblum 1982). The plasmid was also shown to replicate successfully in several Gram-positive and Gram-negative species. Thus, this plasmid contained all the components required for replication and selection in *S. carnosus*. Another advantage of this plasmid was the fact that all of its sequence was readily available in the NCBI database. pC194 is not commercially available, and consequently a variant containing most of its backbone sequence, pCT20, was used instead for the purposes of this project.

The vector pTTQ18 was used as the source for the Gram-negative components of the design. This vector is a variant of the pUC18 containing a *trip-lac* promoter (pTac) upstream of a MCS, together the lac regulation represser gene *Lac*I (Stark et al. 1987). Most of the backbone of this vector was used as part of the novel expression construct for a number of reasons. Firstly, the pUC vector series have been used extensively for cloning purposes, and therefore have already been sequenced and characterized as cloning systems in *E.coli* hosts. Consequently, pTTQ18 had already been proven as a high-copy number cloning vector that uses a derivative of the ColEI origin of replication from pBR322 and can be selected using ampicillin as the selective antibiotic. The sequence of this vector was also readily available on NCBI. Secondly, the pTac promoter and accompanying regulatory system could be used for the purpose of expressing biocatalysts in *E. coli* by replacing the MCS of pTTQ18 for the synthetic linker, which itself contained a wider range of unique restriction sites. The use of the promoter from the Gram-negative components of the novel vector design went against the original approach, in which both Gram-positive and Gram-negative promoters and respective regulatory systems would be intrinsic to the synthetic linker. However, this change in the approach greatly decreased the complexity and length of the linker.

In turn, the synthetic linker was designed around the restriction maps of the other components. Thus it included restriction sites at the extremities that allowed for the assembly of the pC194 and pQTT18, while also containing a MCS with unique sites that were not duplicated in the backbones of these vectors. The MCS region of the linker was designed to fit downstream from the pTac promoter of pTTQ18 and from a second, Gram-positive promoter that was part of the synthetic linkers. Unique restriction sites were also designed to flank the ends of this second promoter so it could also be re-arranged or replaced to fit the needs for biocatalyst expression.

4.2- The Modular Nature of the Promoter Region

The synthetic linker was designed as the centerpiece of the novel expression vector design. Accordingly, this component was designed to work as a linker that allowed for the joining of the other components, while also providing the ability to clone and express heterologous proteins in both Gram-negative and Gram-positive bacteria. Being the only synthetic part of the vector, i.e. constructed from scratch, allowed for a greater freedom in arranging the restriction map of the promoter region to fit these purposes. Figure 31 shows the representation of the synthetic linker as it was designed, and highlights the role of the restriction sites as the flexible hinges around which the other gram-positive and gram-negative components are reversibly connected.

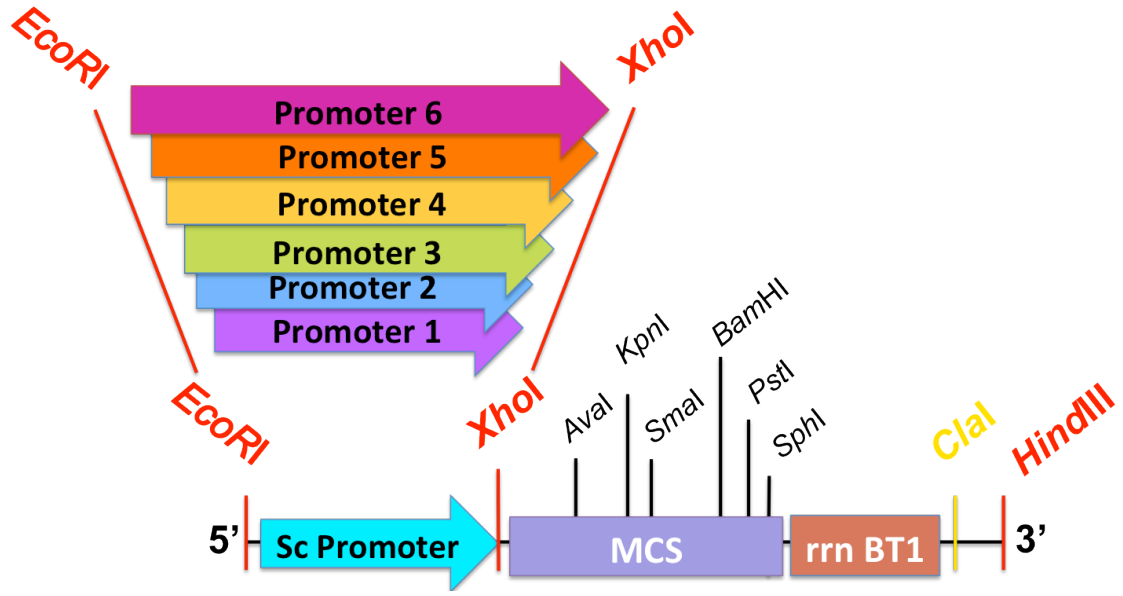


Figure 31- Design for the synthetic linker. The promoter region is flanked by unique *Eco*RI and *Xho*I sites that allow for the exchange between different promoters depending on the needs of the expression system. The MCS contains a set of unique enzymes that allow the cloning of genes encoding heterologous proteins, and is followed by the intrinsic transcription terminator rrn BT1. The 3' end of the synthetic linker contains restriction sites for the joining of the other vector components.

One of the crucial concepts behind the design of the synthetic linker was this idea that it worked as the joining point between the backbone of pTTQ18 and part of pC194. The easiest way to guarantee this was to introduce restriction sites on the linker that aligned to restriction sites in each of the other components. Thus, the ends of the synthetic linker were designed to contain the restriction sites *Eco*RI and *Hind*III, which allowed for the replacement of the pTTQ18 MCS. In addition, the restriction site *Cla*I was introduced after the MCS to allow for the insertion of the pC194 components into the pTTQ18-synthetic linker ligated backbone.

The decision to design the first step for the construction of the novel expression vector around the insertion of the synthetic linker into pTTQ18 was based on two reasons. Firstly, by using pTTQ18 as the starting point for the cloning strategy would allow for the use of *E.coli* as the cloning strain, which had been already extensively characterized as a host for genetic cloning processes and a large array of tools have been devised to facilitate cloning into the latter. By contrast, cloning and transforming *S. carnosus* has been shown to be a much less efficient process (Löfblom et al. 2007), and therefore relying on it as the primary cloning strain would not be ideal.

Secondly, pTTQ18 already contained a MCS that was redundant for the purposes of the vector construction. Thus the synthetic linker was designed with a MCS segment that would replace the latter as the main cloning region for heterologous proteins. Thus, the restriction sites in the end of the synthetic promoter region were complementary to sites in pTTQ18 located upstream and downstream of the MCS, but still downstream from the pTac promoter of the vector to allow *E.coli* expression of the heterologous proteins.

The MCS segment of the synthetic linker was itself designed as a cloning platform for heterologous proteins. As a result, the restriction sites organized within this segment were chosen from a bank of commercial endonucleases that are readily available and consequently would be commonly used in cloning processes. The choice of unique restriction sites was still dependent upon the restriction maps of the other vector components, which had to be checked for homologous sites that would duplicate those in the MCS region. In addition, the MCS was designed with

the cloning of the CHMO gene from pQR239 in mind, and therefore *Xho*I and *Bam*HI was added specifically for this purpose.

The final crucial component of the synthetic promoter region that allowed the expression of proteins in *S. carnosus* was the Gram-positive promoter sequence, which was designed to be located upstream of the MCS and downstream from the pTac promoter of pTTQ18. As with every other component of the synthetic cassette, the promoter segment was also designed with restriction enzymes flanking its borders that would allow for its replacement with other promoter sequences at will. This flexibility to replace promoters at will is the feature that makes this vector design novel. No previously reported shuttle vectors for protein expression in *S. carnosus* had been designed to allow the creation of promoter libraries for *S. carnosus*. In the case of this project, this feature was particularly important because it meant the expression of the biocatalyst would not be irreversibly fixed to one promoter, and could be modulated depending on the needs of the biocatalytic process. It would also allow us to study novel constitutive and regulated expression systems for *S. carnosus*.

4.3- Scanning *S. carnosus* Genome for the Putative Promoters

Expression of heterologous proteins in the novel vector construct was designed around the activity of two distinct promoters: the Gram-negative pTac inducible promoter from pTTQ18, and a Gram-positive promoter that would allow for strong protein expression in *S. carnosus*. While there is a solid database available for well-characterized, commercially used promoters for *E. coli*, the same cannot be said for the staphylococcal bacterium. Therefore, the choice of promoter for the synthetic linker was not a straightforward task.

There are only a couple of staphylococcal promoters that have been characterized in the context of heterologous protein expression. One of these is the lipase promoter (Götz et al. 1985), which was originally cloned from

Staphylococcus hiycus and was used in the vector pNW21 to expressed heterologous protein constitutively in *S. carnosus*. Alternately, the inducible xylose promoter (Wieland et al. 1995), originally from *Staphylococcus xylosus*, was also used routinely in staphylococcal expression systems. The inducible nature of this promoter involves repressor XylR gene, which is co-expressed in the expression vector system and inhibits promoter activity in the absence of xylose.

One valid approach for the choice of promoter would be to use one of these characterized promoters for the expression of heterologous proteins. However, this option was discarded in favor of the choice to use a promoter that was taken *de-novo* from the *S. carnosus* genome. One reason that drove this decision was the fact that none of the promoters listed above are native from *S. carnosus*, and while they have been extensively characterized, they are not guaranteed to be optimized systems for this bacterium. Alternately, there is an high probability promoters native to *S.carnosus* are more optimized for intracellular protein expression in the latter. The other reason for this choice was the discovery that comes from the screening of the Staphylococcal genome for novel promoters. Identifying novel promoters that could work as units for efficient expression in *S. carnosus* would expand the list of functional promoters for staphylococcal species, while also increasing the general understanding of the staphylococcal genome's expression and regulatory systems.

However there was also a greater inherent risk in adopting this more open-ended approach. In contrast with many other bacterial strains, there is very little available information regarding the transcriptome of *S. carnosus*. Additionally, there are very few guidelines as to what genetic elements constitute a strong promoter in this bacterial strain, or how they are regulated. Indeed, many studies have shown that Gram-positive promoters have an high degree of variation from the consensus sequences. Consequently, the selection of promoters from the genome of *S. carnosus* was hampered by the fact that they could not be easily identified, and there was no direct means of testing for their activity and regulation.

Therefore, focusing on the identification of specific promoter sequences in the genome of *S. carnosus* would be counter-productive, as there would be no way to know how efficient these sequences were or if they were natively regulated.

Instead, the methodology adopted in this study for the selection of native staphylococcal promoters was based on the use of *in-silico* predictive tools to correlate native protein expression to promoter strength, the assumption being that highly expressed proteins have an higher chance of being under the influence of a strong promoter. There were several advantages to this approach. Firstly, there are several open-source software available for the prediction of highly expressed proteins that use sequence data as the only criteria. In addition, several studies have been done on the proteomics of other pathogenic staphylococcal species that share an homologous group of core genes (Rosenstein & Götz 2010), and consequently data on these genes could be used to inform predictions of protein expression in *S. carnosus*.

Figure 32 shows the graphical representation of this methodology. The genome of S. carnosus was initially screened for genes that had been identified as highly expressed proteins in the proteomics studies with other staphylococcal species. Predictive software was then used on these genes to predict their expression in *S. carnosus*. A library of potential native promoters was subsequently constructed from the DNA sequences upstream of predicted highly expressed genes.

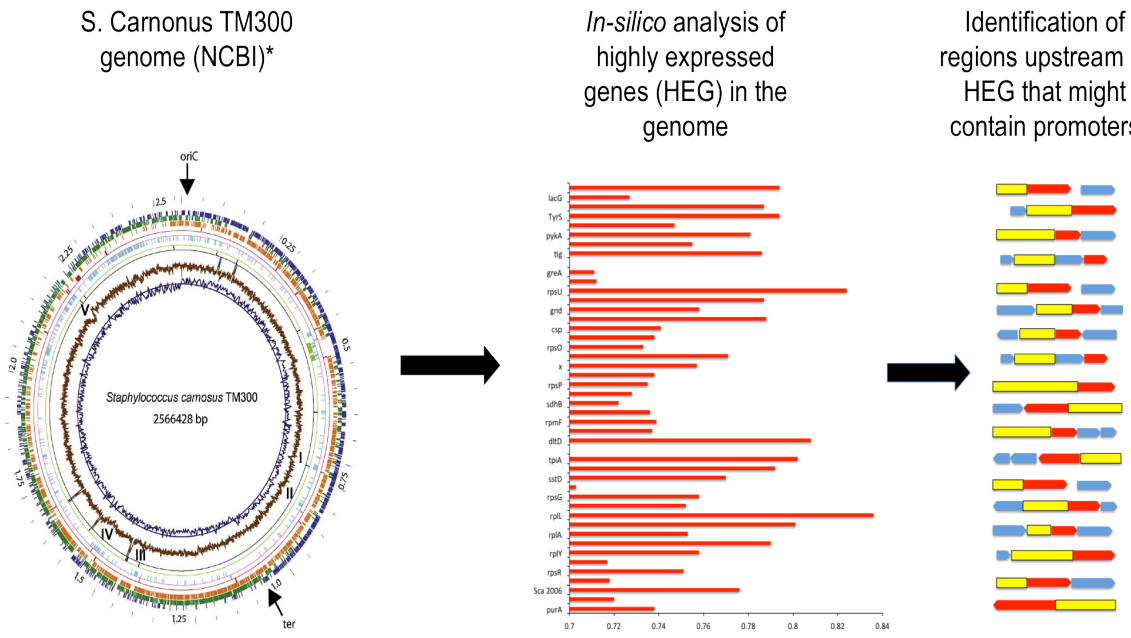


Figure 32- Screening strategy for the identification of putative promoters from the genome of *S. carnosus*. Potential HEGs are identified via a combination of *in-silico* approaches, and regions upstream of these genes are highlighted as hotspots for strong promoters (yellow boxes). *- (Reproduced from the *International Journal of Medical Microbiology*, vol. 300 (2-3), Rosenstein & Gotz, Genomic differences between the food-grade S. carnosus and pathogenic staphylococcal species, pg. 104-108, Copyright © 2010, with permission from Elsevier.)

The predictive software used for the construction of the promoter library was based on the Codon Adaptation Index (CAI) (Puigbò et al. 2008). CAI is a measurement of the codon usage bias of a sequence relative to a reference set, which specifies the resemblance between the sequence and the reference set. Codon bias refers to the observed phenomenon that synonymous codons (that code for a specific amino acid) are not used in equal frequencies in a protein sequence, and this preference to specific codons can be traced to the organism from which the sequence originated (Sharp et al. 2005). This codon bias can be very species-specific, and is viewed as one of the major mechanisms employed by cells to regulate rates of protein expression (Plotkin & Kudla 2010). Thus, there is a correlation between the codon bias of a protein and how well they are expressed. Highly expressed proteins are more likely to have codons which are preferred and consequently in greater abundance in the native organism. Conversely, proteins that contain a high proportion of rare codons, i.e. codons that are not preferred by

the native species, are more likely to have a low level of expression. In fact, several studies have documented the optimization of heterologous protein expression by adapting the codon bias to that of the host organism (Gvritishvili et al. 2010; Kofman et al. 2003).

The software used for the prediction of highly expressed proteins (http://genomes.urv.cat/CAIcal/) calculated the CAI of query sequences as a function of the difference between the codon bias of this sequence and that of a large group of sequences from the genome of the native organism, referred to as the reference set. The values for the CAI vary from 0 to 1, with a linear positive correlation between the value and similarity between query sequences and reference sets. The absolute CAI values are thus inherently dependent on the reference set, and will be interpreted differently when distinct reference sets are used. This is a limitation of the predictive system that warrants caution when choosing the reference set. For instance, when using genomic reference sets, i.e. composed of a broad range of genes from a genome, the CAI value indicates how adapted the query sequence is to the general codon bias of an organism, which is a more indirect measurement of well expressed the gene in the latter. On the other hand, using reference sets composed of genes with specific functions can introduce erroneous biases to the analysis by focusing the CAI on function rather than expression.

There are other limitations to the CAI analysis. On one hand, the CAI can be affected by biases other than the codon bias of the reference set, such as biases in amino acid or nucleotide composition, which creates false positives in the analysis of the data. On the other hand, in organisms that do not display a dominant translational bias, the CAI will be very high (close to 1) regardless of the codon bias of the query sequence, and this can result in a loss of predictive power (Puigbò et al. 2008). To counteract these limitations, the software allows for the calculation of the expected CAI (eCAI) of a sequence. The eCAI is calculated as the average of CAIs from 500 procedurally generated sequences with the same amino acid composition of the query set, but random codon assignments. By randomizing the codon usage of this set, the software is accounting for the statistical noise created by mutations, amino acid or nucleotide biases. Thus eCAI

represents the threshold above which the values of CAI can be statistically attributed to correlations between codon biases instead of being a product of biases that are not related to codon usage.

The reference set used for the purposes of CAI calculations in the predictive analysis of gene expression levels in *S. carnosus* was taken from the Codon Usage Database (http://www.kazusa.or.jp/codon/), which contained the codon percentages of a set of 52 proteins with a broad spectrum of activity. The diversity of the set ensured that we were not introducing functional biases to the analysis but instead looking at a protein population that represented the more generalized codon bias of the *S. carnosus* genome. Figure 33A shows the codon bias generated from this set of proteins, which clearly shows the preference towards certain synonymous codons. This is most prevalent in amino acids like alanine (Ala) and leucine (Leu), which can be coded by 4 and 6 codons respectively but show a dominating preference to a single codon.

For the query set, 46 gene homologs to proteins that had been identified through *in-silico* prediction tools (Jansen & Bussemaker 2003) and empirical proteomics (Kohler et al. 2005; Suzuki et al. 2012) , to be highly expressed genes (HEG) in *S. aureus* were selected from the genome of *S. carnosus*. The list of proteins, together with their CAI in several *S. aureus* strains, can be found in the HEG-database (http://genomes.urv.cat/HEG-DB/) (see Appendix 2). Many of these proteins were housekeeping genes that have been highly conserved within staphylococcal species, and were therefore to have similar expression profiles between *S. aureus* and *S. carnosus*. Figure 33B shows the ratios between the CAI for each protein in the query set and the eCAI of the set, obtained through the CAI calculation software. Values above 1 indicate proteins with statistically significant CAIs.

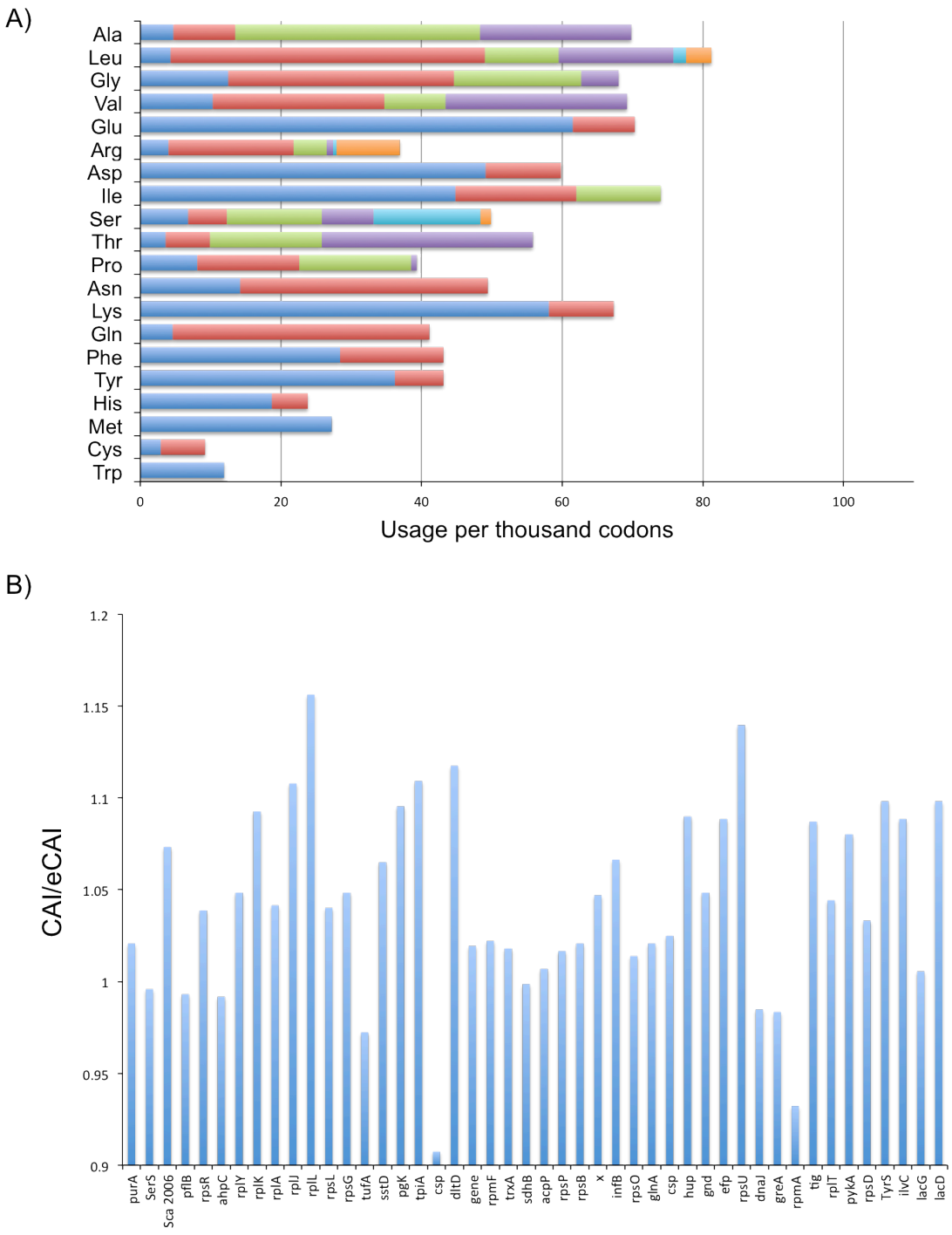


Figure 13- A) Codon usage frequencies of the set of 51 protein sequences taken from the Codon Usage Database. The different colors represent different synonymous codons that translate specific amino acids. A strong codon preference can be observed for amino acids with highest frequencies. B) CAI/eCAI ratio of 46 proteins used as the query set for the CAI analysis. Values above 1 refer to proteins whose CAI is statistically significant.

Most of the proteins in the query set had CAi values above the eCAI value of 0.723, and thus shared more than 70% codon bias with that of the reference set.

Many of the highest CAI values corresponded to ribosomal subunits that, as house-keeping proteins, are expected to express strongly in active cells. In fact, the 50S ribosomal protein L7/L12, which generated to highest CAI in the *in-silico* analysis, had been previously identified as the ribosomal protein with the highest level of expression in both the exponential and stationary phase of *S. aureus* COL (Kohler et al. 2005). The similarities between the results from the CAI analysis and the proteomics data for *S. aureus* suggested that the CAI values could be directly related to protein expression levels, and consequently could be used as reasonable predictive tools for HEG.

Once potential HEG were identified, the sequences immediately upstream of the genes were selected as potential strong promoters (Figure 34). In some cases, these sequences were very short, or were part of a different gene, suggesting that either the promoter for the selected HEG was contained within another gene, or that the HEG was part of an operon. Operons can be under the influence of strong promoters, but in many cases this activity is also strongly regulated. Since the initial focus of the promoter selection approach was to select for promoters that could induce a high level of constitutive protein expression, regions upstream of putative operons were discarded because they could contain elements of unknown regulatory activity. As a result, only sequences upstream of single genes were chosen.

One such gene was the 50S ribosomal protein L10 (*rplK*). The sequence immediately upstream was a stretch of 183 base pairs that contained several promoter-like DNA motifs, suggesting that it could contain multiple promoter activities (Figure 35). Considering the potential for high activity due to the multiplicity of promoter elements, and the assumption that ribosomal proteins are more likely to be constitutively expressed throughout the bacteria cell cycle, this short sequence was selected as the starting Gram-positive promoter, designated the rplK promoter, to be included as part of the synthetic linker.

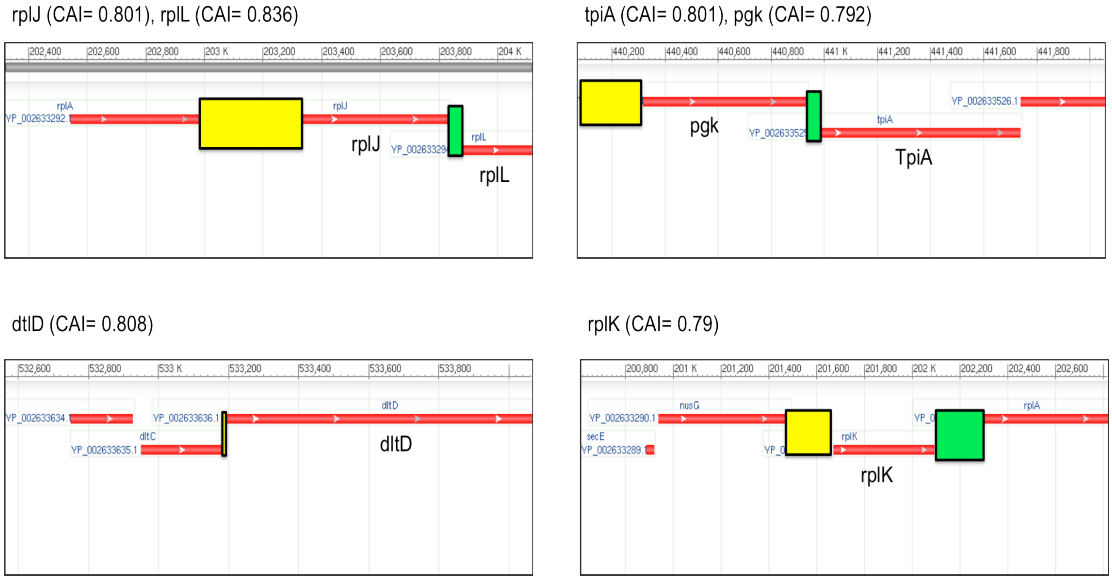


Figure 14- Position and orientation of genes with high CAI values in the genome of *S. carnosus*. The sequences upstream of the genes and the spacer sequences between genes are represented by the yellow and green boxes, respectively. In cases like the gene tpiA or dltD, which generated the highest CAI, the space between these and adjacent genes is too small to accommodate for promoters (green boxes), which in turn suggests that the genes are part of an operon. This is not the case for the rplK gene, as sequences upstream and downstream for it are long enough to consider this gene as an individual expression unit, not associated with the expression of other genes.

- Putative rplK promoter

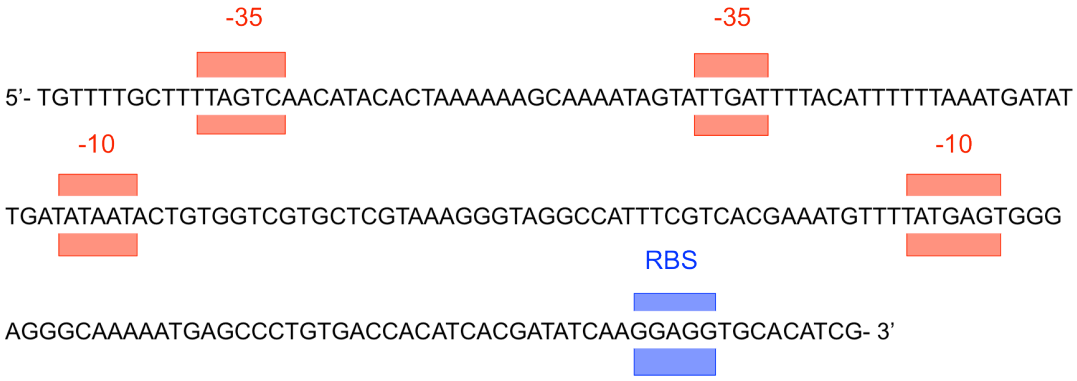


Figure 35- Sequence upstream from the rplK gene. Regions with putative -35 and -10 promoter elements are highlighted within red boxes. The putative ribosomal binding site is shown within the blue box.

4.4 – The Construction of the pQR1029 Vector

The novel expression vector, designated pQR1029, was created in two steps as outlined in Figure 36. The first was the insertion of the synthetic linker into the backbone of pTTQ18, thereby replacing its MCS and adding the required restriction sites for the insertion of the Gram-positive fragment. The second and final step was the ligation between the product of the pTTQ18/ synthetic linker ligation and a fragment from pCT20 containing the chloramphenicol resistance gene (CAM), and the origin of replication for staphylococcal strains.

For the first step, both pTTQ18 and the synthetic linker were restricted with *Eco*RI and *Hind*III. The linker was created de-novo by the company DNA 2.0 and provided in a cloning plasmid with the same flanking restriction sites, so that it could be extracted prior to ligation by using the same digest. The restricted fragments were subsequently purified through gel extraction and ligated with T4 ligase from NEB. After ligation, the resulting DNA products were transformed into chemically competent *E. coli* TOP10 and selected using ampicillin selective plates. Colonies were subsequently screened for positive ligation through restriction digests.

Being supplied by a commercial company, the synthetic cassette had a high degree of purity, and therefore the ligation protocol did not require fine-tuning and trouble-shooting to accommodate for sub-optimal conditions. In fact, only one set of screening assays was required to identify colonies containing the positive ligation.

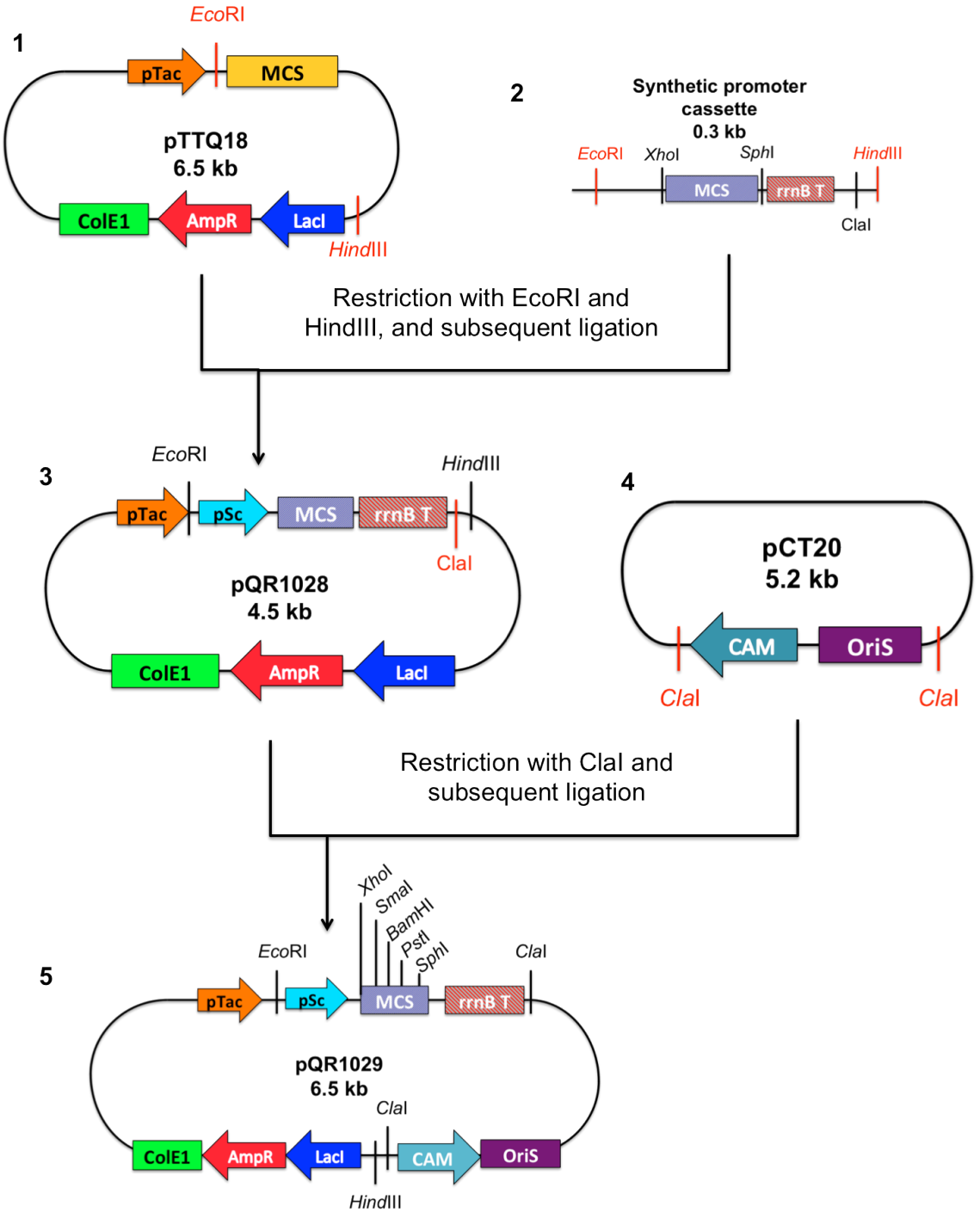


Figure 36- Ligation strategy employed for the construction of the vector pQR1029. The first step was the creation of the intermediate pQR1028 (3), and involved the removal of a 2.3 kb region from pTTQ18 (1) via endonuclease restriction with *Eco*RI and *Hind*III, and subsequent ligation of the remaining backbone with the synthetic linker (2), which had been extracted from a commercial vector using the same restriction enzymes. In the second step, the pCT20 fragment containing the CAM resistance gene and the origin of replication is digested from pCT20 (4) and ligation with pQR1028, creating the shuttle vector pQR1030 (5).

Figure 37 shows an agarose gel with the screening results from the first ligation step. The plasmid DNA was extracted from colonies that had been grown overnight in ampicillin selection media and subsequently digested with *Xho*I. The insertion of the synthetic linker into pTTQ18 had replaced the MCS of the vector with a set of unique restriction sites, including *Xho*I, that are not present in re-ligated or undigested pTTQ18. As a result, only the ligation product of pTTQ18 with synthetic linker would be restricted with *Xho*I, resulting in a visible size difference between cut and uncut vector in the agarose gel (lanes 12, and 13), while the re-ligated vector would remain uncut (lanes 2 and 3). Sequencing of the vector around the pTac promoter confirmed the results from the screening assays (Appendix 1).

The ligation construct, pQR1028, contained the restrictions sites from the synthetic linker that were required for the second ligation step.

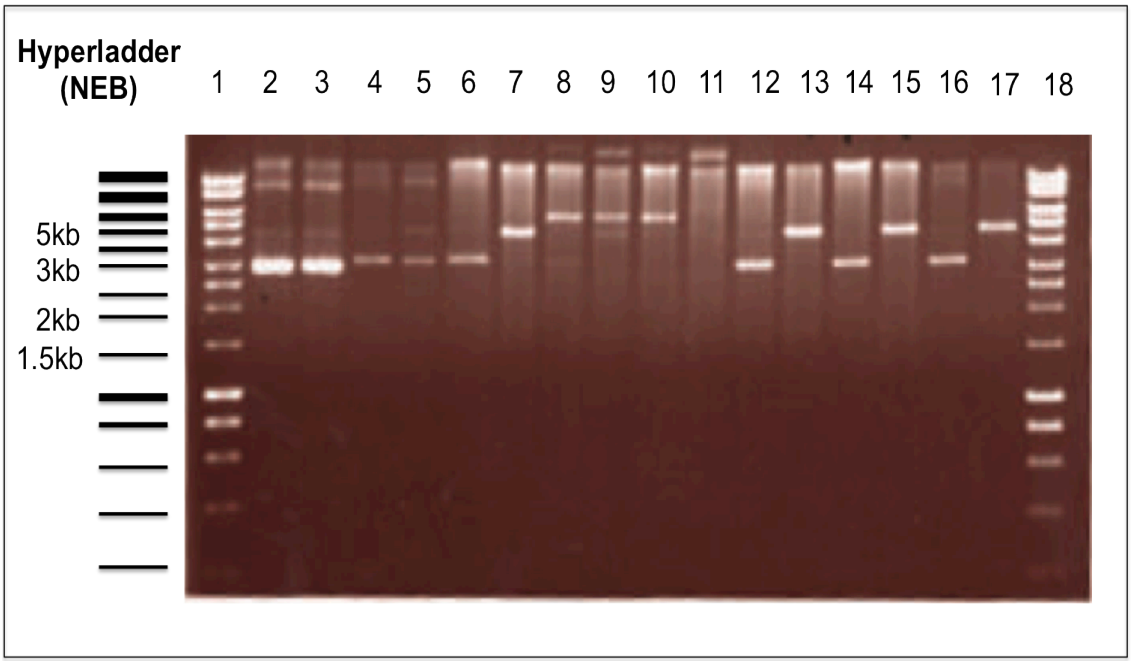


Figure 37- Agarose gel (0.7% w/v) with the results from the screening of the pQR1028 construct. Lanes 1 and 18: Hyperladder (NEB); lanes 2 to 17: 6 plasmid samples from *E.coli* colonies selected on ampicillin plates in uncut and cut states, respectively. Lanes 12 to 17 showed the profiles we were expecting to generate when cutting the plasmid pQR1028 with *Xho*I. Indeed, cutting circular plasmid (lanes 12, 14, 16) with this enzyme produced a linearized band about 4.5 kb in size (lanes 13, 15, 17).

The second step in the construction of the shuttle vector pQR1029 involved the restriction of both pCT20 and pQR1028 using the endonuclease *Cla*I, followed by ligation of the resulting pCT20 fragment containing the origin of replication and the CAM gene with the linearized vector.

Unfortunately, this step was not as straightforward as the first. Indeed, the fact that the pCT20 vector ha been provided inside *S. carnosus* TM300 introduced the problem of having to extract the plasmid DNA from the staphylococcal strain. Due to the toughness of the peptidoglycan cell wall, extraction protocols for Gram-positive bacteria are work intensive and often result in very low product yields. In addition, these protocols require the use of expensive lytic enzymes such as lysostaphin and lysozyme to break up the cell wall. Finally, the DNA extraction from the rest of the cell debris involves the use of a phenol-chloroform two-phase system, which is not only extremely toxic at higher volumes, but also results in low purity DNA samples. These protocols are therefore limited to small-scale extractions, and are not optimized to achieve highly concentrated and pure DNA products.

Consequently, the initial attempts of extracting pCT20 from S*. carnosus* using these conventional extraction protocols produced unreliable results, with large fluctuations of plasmid yield that primarily depended on the fine handling of the phenol-chloroform gradient. Most of the extracted samples also contained a large quantity of ribonucleic acid impurities, making them very hard to visualize on an agarose gel. As a result, the samples obtained from these extraction protocols were not pure or concentrated enough to allow for efficient restriction and ligation.

The low concentration of plasmid DNA in the samples was the main bottleneck for the ligation, due to the fact that most of the plasmid would be lost during the required purification of restricted fragments from an agarose gel. As gel extraction protocol generally produce low yields of purified product, starting with a low concentration of plasmid before restriction and gel purification would result in very low or negligible final plasmid concentration.

As a response to this bottle-neck, two solutions were devised to ensure plasmid yields: the first was to develop a reproducible and less complicated method for extraction of the plasmid DNA from large samples of *S. carnosus*; the second was to directly purify high quantities of the desired fragment from pCT20

contained within un-lysed samples by using PCR amplification with primers that flanked the fragment of interest.

(i) The plasmid extraction protocol

Figure 38 shows a schematic of the type of factors that were considered when developing a more efficient protocol for plasmid extraction from the staphylococcal strain. The first stage in the protocol would be the disruption of cells. Previously described protocols used a mix of lysogenic enzymes to lyse the thick peptidoglycan and lipid layers of *S. carnosus*. Therefore the first line of enquiry was whether there was a quicker and cheaper disruption method that could produce high level of lysed material. One alternative would be to use physical methods to induce disruption of the cells, such as sonication or French-press induced homogenization. All the results using physical disruption methods indicated that some degree of chemical treatment was required to weaken the peptidoglycan layer of the cells for the mechanical disruption to be efficient.

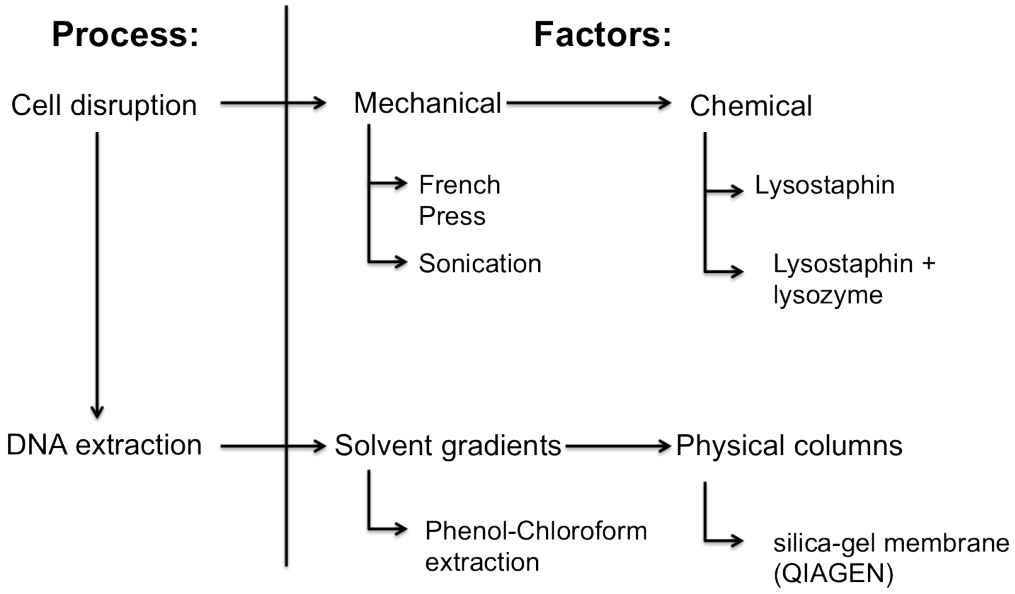
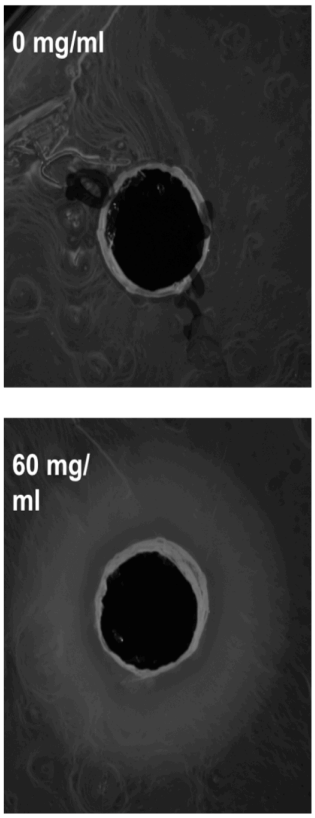


Figure 38- Factors that were tested on *S. carnosus* for the extraction of plasmid DNA.

Several staphylococcal strains, including *S. aureus* and *S. epidermis*, have been reported to be resistant to lysozyme-mediated cell disruption, and thus addition of lysostaphyn is needed to promote efficient lysis (Wu et al. 2003). To test if the same was true for the genetically related *S. carnosus*, toxicity assays were conducted in which samples of *S. carnosus* were grown overnight on non-selective plates containing various concentrations of lysozyme. The results from these assays (Figure 39 A) showed that concentrations of lysozyme as high as 60 mg/ml did not seem to inhibit growth of *S. carnosus*, from which it could be concluded that the peptidoglycan wall was not sensitive to lysozyme digestion.

The inefficiency of this enzyme to degrade the peptidoglycan wall posed the question of whether the presence of lysozyme in a lytic cocktail was really necessary. To access this, the effect of lysozyme on lysostaphin-mediated cell lysis in different buffers under different incubation condition was tested. Bacterial samples were grown overnight, pelleted, and resuspended in various lytic solutions containing a constant concentration of lysostaphin but different concentrations of lysozyme and different reaction buffers. The resuspended samples were subsequently incubated at 37 ^{0}C for one and two hours, after which a small volume of the resulting lysates was loaded unto an agarose gel to check for the release of plasmid DNA from lysed cells. The results (Figure 39A and 39B) showed that lysostaphin by itself was sufficient to promote efficient cell disruption. In fact, adding high concentration of lysozyme to the samples seemed to have a detrimental effect on the quality of DNA extracted from the lysed samples (figure 39C). It is uncertain why samples digested with both lytic enzymes did not show visible bands on the agorose gel, since these samples were still lyses by through the activity of lysostaphin. One hypothesis could be that lysozyme binds to the plasmid DNA, retaining it in the well of the agarose gel and consequently preventing it from migrating into the polymer matrix. One way to test this would be to heat-inactivate the lysis before loading the DNA onto the gel, which should prevent the DNA from being sequestered by active enzyme.

A)



B)

Sample	Reaction buffer pH	Lysosyme concentration (mg/ml)	Lysostaphin concentration (µg/ml)	Incubation time (hours)	Phenotype
A	7.4	0	1	1	Plasmid
B	9	0	1	1	Plasmid
C	7.4	60	1	1	No extraction
D	9	60	1	1	No extraction
E	7.4	0	1	2	Plasmid
F	9	0	1	2	Plasmid
G	7.4	60	1	2	No extraction
H	9	60	1	2	No extraction

C)

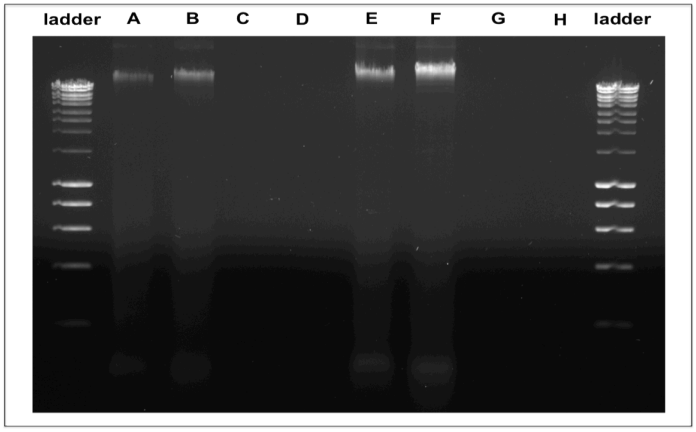


Figure 39- Different test run on the effects of chemical lysis of *S. carnosus*. A) Effect of different concentrations of lysozyme on layers of *S. carnosus* spread on agar plates. Holes were drilled in plate sections after incubation of bacteria on plates, and various concentrations of lysozyme solution were loaded unto these reservoirs and incubated overnight. B) Table specifying an experimental design with the different factors studied for chemical lysis of *S. carnosus* cells. All the factors were studied at two levels, except for lysostaphin concentration, which remained constant throughout. C) Agarose gel (0.7% w/v) containing the results from the chemical lysis experiment. The letters in the lanes correspond to the letters outlined in the experimental design table (B).

The other part of the extraction protocol that needed to be addressed was the purification step. As mentioned above, in the standard protocol this was achieved by applying solvent gradients that required delicate handling and involved the use of hazardous phenol-chloroform based compounds. On the other hand, easy-to-use kits have been developed for plasmid DNA extraction in gram-negative bacteria that applied a solid phase chromatography column technology to bind and purify DNA using vacuums or the centrifugal force of a bench-top centrifuge. The columns were designed to work on standard lab equipment, and had the added

advantage of being optimized for plasmid extraction. Therefore, the possibility of using these columns for purification of plasmid DNA from lysates of *S. carnosus* was tested. Samples that were either mechanically or chemically lysed were centrifuged, and the supernatant was subsequently passed through the solid phase extraction colonies. This method resulted in samples containing pure plasmid DNA at high concentrations.

Thus, after identifying the factors that were crucial to DNA extraction from *S. carnosus*, an optimized plasmid extraction protocol was developed that did not require complicated enzyme cocktails and took advantage of the ability of gram-negative purification kits to obtain highly pure DNA plasmids from bacterial samples. In addition, this optimized protocol could easily be adapted for extraction of large volumes of cell lysate. The full protocol is described in the materials and methods chapter (see chapter 2, section 2.2.1).

(ii) PCR amplification of pCT20 fragments

The creation of a protocol that could reliably extract the plasmid pCT20 from *S. carnosus* only solved part of the problem. The restriction of the pCT20 and subsequent purification of the fragment bearing the replication and selective components still required running the sample on agarose gel to separate the different sized fragments and purify the desired components via gel extraction. This method is very inefficient and often results in very low concentrations of the end product, which in turn decrease the probability of ligation to occur. The strategy devised to work around this problem was based on the amplification of the desired fragments from pCT20 by using polymerase chain reaction (PCR). The direct amplification of the desired fragment from pCT20 not only took away the need to use gel extraction for purification of the correct fragment, but also had the added advantage of exponentially increasing the final concentration of the fragment (Figure 40). The primers used for amplification were designed to anneal 20 base pair (bp) sequences flanking both the origin or replication and antibiotic resistance gene from pCT20, and also contained overhangs with the restriction sites for *Cla*I and *Hind*III to allow for ligation with pCT20.

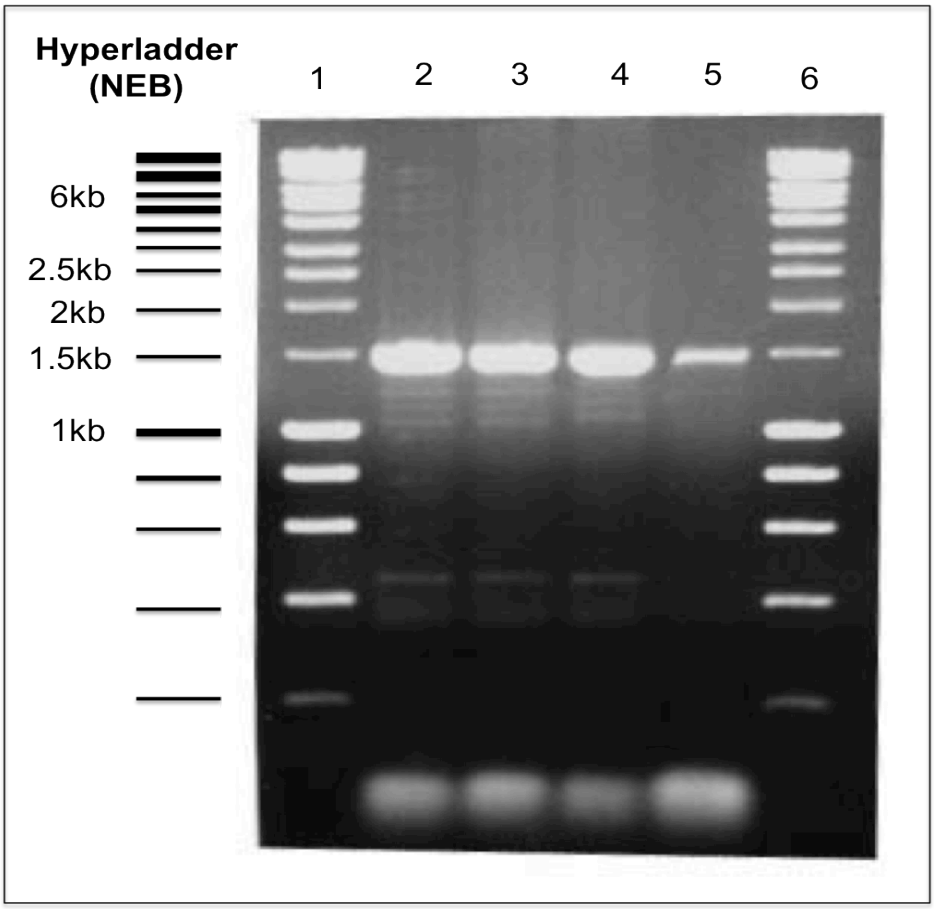


Figure 40- Agarose gel (0.7% w/v) containing the products from the pCT20 PCR amplification. The expected size for amplified fragment of pCT20 containing CAM and oriS is 1.5 kb. Lanes 1 and 6: hyperladder (NEB); lanes 2 to 5: PCR reactions using different primer annealing temperatures (64.5, 65.5, 66.4, 67.9, respectively).

Thus, a methodology combining the optimized plasmid DNA extraction protocol with subsequent amplification using sequence specific PCR resulted in high final concentrations of pure pCT20 fragments containing the genetic elements of interest.

Subsequent attempts to ligate these the amplified fragments from pCT20 with the linearized pQR1028 were still unsuccessful due to a second bottleneck that had not been accounted for when designing the vector ligation strategy. This bottleneck was the effect of DNA methylation of the restriction of specific sites by endonucleases. Methylation is a mechanism by which cells regulate several genetic processes, including genome replication and protein expression. Methylation also shields DNA from digestion by intracellular endonucleases, a

mechanism that selects for the degradation of foreign sequences that are not methylated in the same manner. One endonuclease that are susceptible to inhibition by methylation is *Cla*I. This posed a problem for the second and final step in the pQR1029 construction process, whereby ligation was dependent on a *Cla*I that had been specifically designed as part of the synthetic linker. Since the *Cla*I site on the ends of the PCR amplified product from pCT20 would not be affected by methylation, since it was not subjected to the cellular methylation process, the same site on pQR1028, which had been replicated in *E. coli*, was shield from restriction with the corresponding endonuclease. In fact, the *Cla*I site in pQR1028 (ATCGAT) overlapped with a recognition site for the methylation protein DNA adenine methyltransferase (Dam), GATC (Figure 41). As a result, the *Cla*I site of pQR1028 was protected and could not be cut in the plasmid extracted from the *E.coli* TOP 10 strain, which had been routinely used as the cloning strain. To solve this problem, a Dam-/Dcm- *E.coli* strain, in which the genes for the methylation pathway have been knocked-out, was used as the cloning strain for pQR1028 replication and extraction. Switching to this strain increased the risk of DNA alteration, since non-methylated DNA is unprotected and more liable to damage, but the risk was necessary for the ligation to be possible.

After solving all inherent problems with the second ligation step, the vector pqr1029 was successfully ligated. The ligations results were confirmed by sequencing of sequences spamming pQR1028 and one of the ends of the pCT20 fragment. The vector was also shown to grant resistance to both ampicillin and chloramphenicol in *E. coli* TOP10, indicating that the selectivity components of the vector were working as intended.

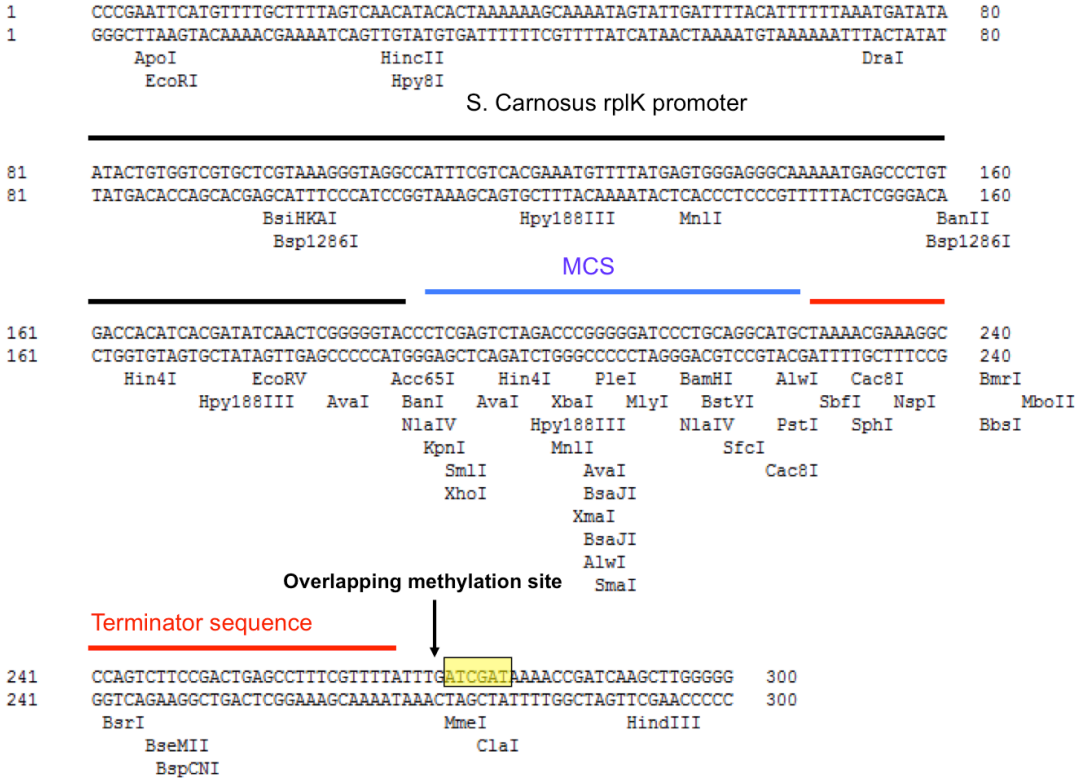


Figure 41- Sequence for the synthetic promoter linker, in which the different elements are highlighted with different colored lines: black line: rplK promoter; blue line: multi-cloning site (MCS); red line: the hairpin terminator sequence. The yellow box refers to the *Cla*I restriction site, which overlapped with a methylation site (black arrow).

4.5- Ligation of the CHMO Gene to pQR1029

The construction of the shuttle vector pQR1029 was a milestone in establishing an expression system for both Gram-negative and Gram-positive bacteria. In the context of the project, an additional ligation step between the CHMO gene and pQR1029 was required to allow expression of the biocatalyst inside *E. coli* and *S. carnosus*. Accordingly, the MCS of the vector had been designed to permit the direct transfer of the gene sequence from pQR239 via a *Xho*I-*Bam*HI restriction.

This ligation was also more problematic than expected, and a variety of ligation strategies were tested as solutions for possible problems. The standard ligation protocol resulted in a large number of transformed colonies, but all of those that were screened showed re-ligation of pQR1029. The restriction of the vector with both enzymes *Xho*I and *BamH*I was very hard to access, since the restriction sites were located very close together and resulted in a non-detectable decrease in the size of the linearized vector. Thus, the high rates of re-ligation suggested that while the vector was being linearized, one of the restriction sites could be working improperly. To test this, restrictions of pQR1029 with each endonuclease individually were performed, and the results from these tests confirmed that both enzymes were digesting the DNA. Therefore it was concluded that the large percentage of re-ligations could not be explained by a sub-optimal restriction by the endonucleases.

Consequently, as a preventive measure to inhibit re-ligation of the vector pQR1029 before ligation, the enzyme Antarctic phosphatase (NEB) was used. This enzyme catalyzes the removal of 5' phosphate groups from the exposed ends of double-stranded DNA, preventing self-ligation of the strands. Samples treated in this manner did not produce any colonies on selective agar plates, even after inactivation of the phosphatase through heat treatment of samples at 65 ^{0}C for 10 minutes. On one hand, the results confirmed that the large number of transformed colonies seen in the first set of ligations was due to re-ligation of the vector pQR1029. On the other hand, these results suggest that either the Antarctic phosphatase was not efficiently inactivated and interfered with ligation, or that the CHMO fragment was not being correctly restricted from pQR239. The fact that no colonies could also be seen after the incubation time of heat treatment of Antarctic phosphatase-treated samples was increased from 10 to 20 minutes suggested that the problem lied of the restriction of CHMO from pQR239.

In the conventional extraction method, the CHMO gene was purified from the pQR239 vector via gel extraction after restriction with a *Xho*I/*Bam*HI restriction mix. While the process was quite straightforward, as the size of the gene (1.6 kb) was smaller than the vector (5 kb) and therefore can be easily identified in an agarose gel, the lower concentrations of the end product could have been a source

for the inefficiency of the ligation process. More important, the trouble-shooting experiments conducted of the restriction of pQR239 identified a crucial bottleneck with the restriction itself. In fact, restriction of the vector pQR239 with *Bam*HI produced a similar restriction profile to a double-digest with a *Xho*I/*Bam*HI restriction, which highlighted the presence of a *Bam*HI site upstream as well as downstream of the CHMO. Since the restriction map for pQR239 did not account the presence of a second *Bam*HI site upstream of CHMO, it was impossible to know if the *Xho*I-*BamH*I double digest produced the gene fragment containing the corresponding sticky ends, or a fragment containing *Bam*HI overhangs at both ends of the gene. In the case of the latter, ligation to a *Xho*I-*Bam*HI restricted pQR1029 would be impossible, which would explain the unsuccessful ligation attempts when using Antarctic phosphatase-treated pQR1029.

To work around this problem, it was decided that the best strategy would be amplify the gene directly from pQR239 using primers with overhangs containing restriction sites that would allow for ligation of CHMO with pQR1028 in the correct orientation. Thus, the primer aligning to the 5' end of the gene was designed with a *Xho*I restriction site, while the 3' end primer was designed with the restriction site *Pst*I, which ligated to an homologous site downstream from *Xho*I in the MCS region of pQR1028. Using PCR amplification, a high concentration of the CHMO gene was amplified from pQR239, which had been isolated from overnight cultures of *E. coli* TOP10, and subsequently inserted into the MCS of pQR1029. The resulting ligation product, pQR1030, was screened by restriction analysis using a *Xho*I/*Sph*I reaction mix, which should excise the CHMO from the vector backbone is the ligation was successful (figure 42). Sequencing of the area around the pTac and MCS of the ligation product was also performed to confirm that the gene was inserted in the correct orientation (Appendix 1).

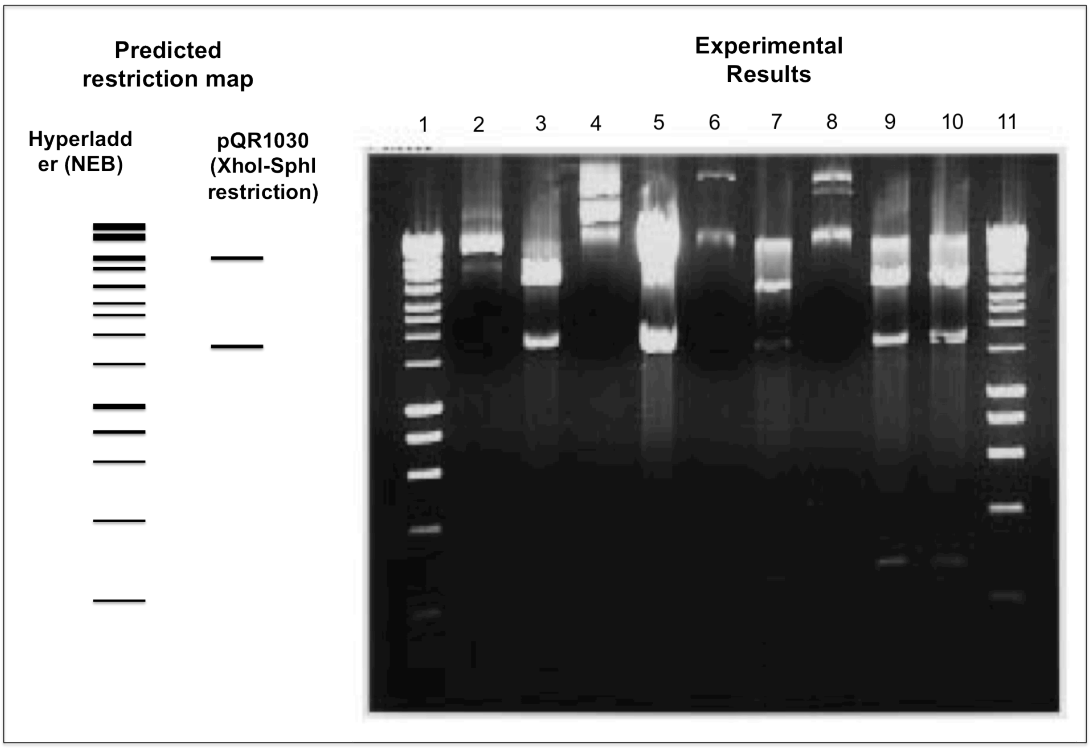


Figure 42- Agarose gel (0.7% w/v) containing the results from the screening of *E.coli* colonies for the pQR1030 construct, together with the expected restriction profile of pQR1030 when digested with *Xho*I and *Sph*I. Lanes 1 and 11: hyperladder (NEB); lanes 2 to 10: plasmid samples from *E.coli* colonies as a circular molecule and cut with *Xho*I and *Sph*I, respectively. The positive constructs show a distinct profile between cut and uncut smaples, with cut samples having to bands corresponding to the linearized pQR1029 (6.5kb) and one corresponding the CHMO gene (1.6kb).

4.6- Initial Biocatalytic Studies with pQR1030 in *E. coli*

The expression vector pQR1030 contained the CHMO gene under the expression of both the Gram-negative pTac promoter and the staphylococcal ribosomal promoter rplK. Theoretically, the set-up would enable the vector to express the biocatalyst in both *E.coli* and *S. carnosus*, provided that the increased distance between pTac and the gene did not impair the expression of this promoter. Since the vector had been cloned into *E.coli*, CHMO expression assays were initially performed on this bacterial strain. The ability to express CHMO in Gram-negative bacteria using pQR1030 would also validate the dual-promoter

vector design as well as trouble-shoot the insertion of the CHMO gene into the pQR1029.

Due to the enzyme's use of the co-factor NADPH during reaction, the level of CHMO expression can be indirectly accessed by measuring co-factor NADPH oxidation rates during biocatalysis. Accordingly, biocatalytic assays were performed on lysates of *E. coli* cultures hosting the vector pQR1030. Lysates were produced from bacterial cultures grown in 250 ml shake flasks containing 50 ml of NB2 medium and 100 µg/ml of ampicillin for 8 hours or 24 hours. ITPG (1mM) was alternately added after 3 hours of incubation to test the induction of the pTac promoter. After incubation, samples were collected and resuspended in 5 ml of Tris-HCl buffer (pH 9), and subsequently lysed using a French-Press. The supernatant from the resulting lysates was collected and subsequently used directly for the NADPH oxidation assays. The operating conditions of the assays did not deviate from previous experiments conducted by O'Sullivan et al. (2001).

NADPH absorbs at a wavelength of 340 nm, which can be measured using a standard spectrophotometer. Thus the biocatalytic reactions were performed in 1 ml-cuvettes by mixing the 0.2 ml of the supernatant from lysates with 0.8 ml of reaction buffer (pH 9) containing 0.161 mM of NADPH and 7.14 g/L of bovine serum albumin (BSA). The resulting mix was inserted in a spectrophotometer with temperature control and incubated for 2 minutes at 30 ^{0}C, during which time the NADPH levels in the sample were measured using a wavelength of 340 nm. This first measurement corresponded to the level of NADPH oxidation from general activity of cellular metabolic enzymes. After this initial period of incubation, 2 mM of the substrate cyclohexanone were added to the sample, and the decrease of the NADPH levels at 340 nm was measured for 2 additional minutes. This decrease corresponded to the level of NADPH oxidation resulting from the combined biocatalytic activity of CHMO and the activity of metabolic proteins also present in the supernatant. The final NADPH oxidation rate from CHMO biocatalysis was subsequently calculated as the difference between the two measured rates.

Figure 43 shows the results from the NADPH oxidation assays. Since the vector pQR239 is the conventional expression system for CHMO expression in *E. coli* TOP10, it was routinely used in the NADPH oxidation assays as a reference point to CHMO expression using pQR1030.

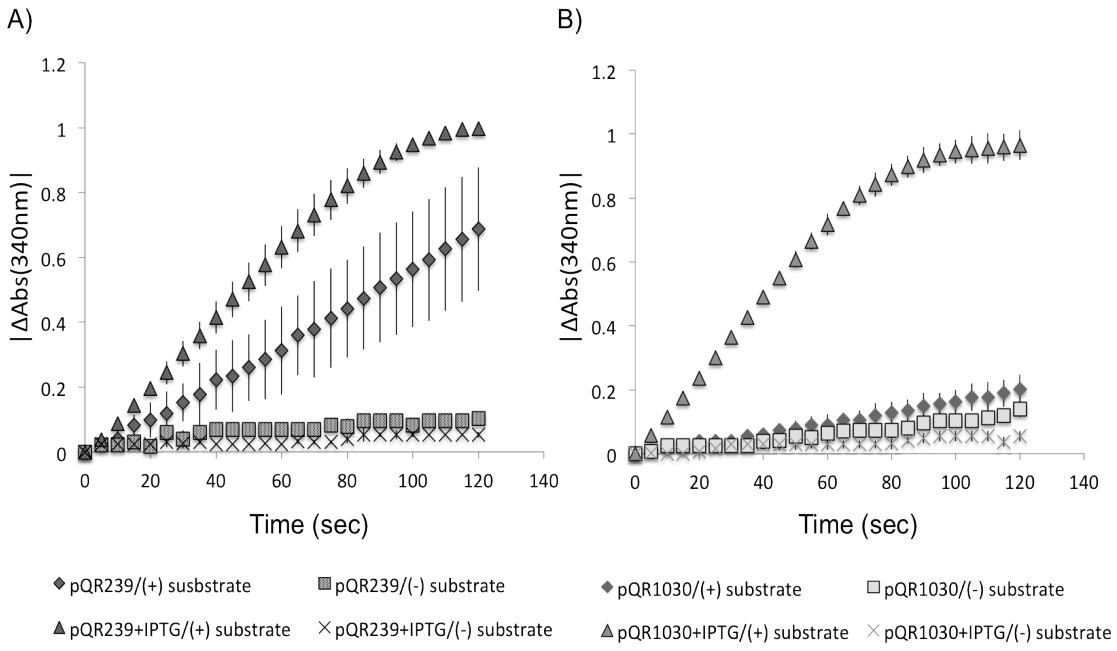


Figure 43- NADPH oxidation profiles of *E. coli* pQR239 (A) and *E. coli* pQR1030 (B) lysates expressing the CHMO gene. NADPH oxidation was detected by measuring absorbance at 340 nm, which decreases as NADPH is oxidized during metabolic activity of CHMO biocatalysis. The values in the y-axis are the absolute numbers of absorbance gradients. The substrate used was cyclohexanone (2 mM), which was added after 2 minutes of reaction in the absence of substrate. IPTG was added to growing cultures at a concentration of 1mM. Error bars refer to the standard deviation of the experimental set-up, calculated from triplicates of the same assay.

Lysates from pQR1030 cultures showed comparable oxidation profiles to those obtained with pQR239 lysates. Both cases exhibited a linear NADPH oxidation rate until shortly after 1 minute of incubation, at which time they start to plateau due to limiting substrate concentration. pQR239 lysates achieved a maximum oxidation rate of 44 µmol of NADPH per minute per gram of dried cell weight (DCW) under IPTG-induced conditions, while the pQR1030 lysates achieved maximal rate of 68.8 µmol min^{-1} g^{-1} (DCW). Alternately, under non-induced condition, pQR1030 samples did not exhibited significantly higher biocatalytic activity when compared with the general metabolic activity in the absence of cyclohexanone, while pQR239 samples still exhibited a linear NADPH oxidation rate of 15 µmol min^{-1} g^{-1} (DCW).

The fact that the NADPH oxidation rates of pQR1030 samples were very low in the absence of IPTG when compared to pQR239 samples grown under the

same conditions suggested that the *Lac*I gene present in pQR1030 was working efficiently as a tight repression system for pTac-mediated CHMO expression. In the case of pQR239, the araC repressor gene is not contained within the expression vector, which would explain the less stringent control of CHMO expression in this system. On the other hand, this result also suggested that the staphylococcal rplK promoter downstream from the pTac promoter did not exhibit any activity in *E. coli*. Therefore, the differences in oxidation rates between pQR1030 and pQR239 was attributed to difference in the expression levels of the pTac and pBAD promoter, with the former showing a higher level of protein expression.

One possible explanation for the lack of activity of the staphylococcal promoter is the presence of AT rich sequences within the latter that could form a hairpin structure upon separation of the DNA into single strands or in the subsequently transcribed RNA (Figure 44). The mechanism of protein expression regulation via RNA secondary structures had been previously suggested for the chloramphenicol resistance gene in Gram-positive bacteria (Lovett 1996), and therefore some promoter regions for S. carnosus could be designed to allow for this regulation to occur. However, if CHMO expression was indeed inhibited during protein translation due to secondary RNA structures, no change in enzyme activity should be seen between IPTG induced and non-induced samples. Consequently, this hypothesis was not considered as a satisfactory explanation for the lack of promoter activity.

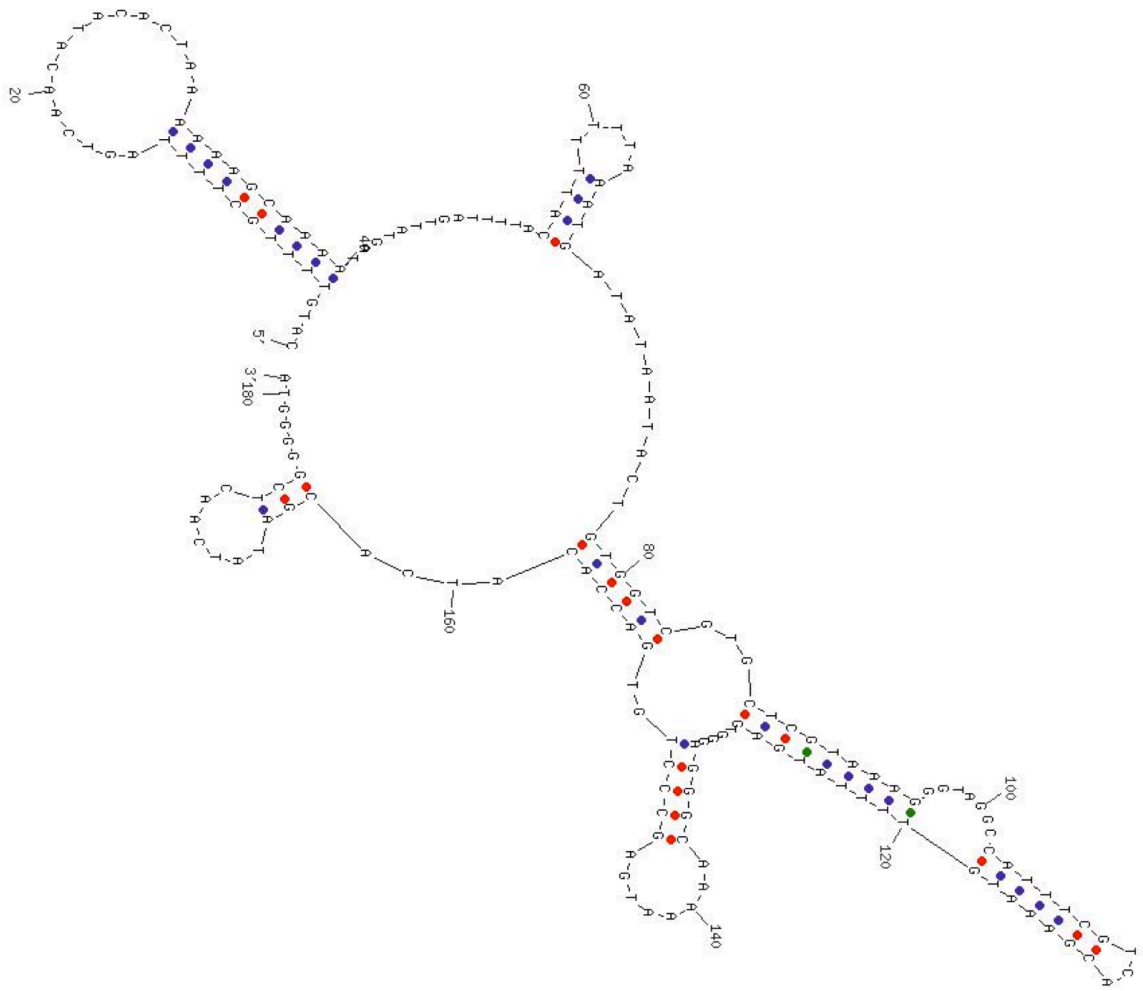


Figure 44- Putative secondary DNA structure of the rplK promoter in the synthetic linker, as calculated via the OligoAnalyzer software (http://eu.idtdna.com/analyzer/Applications/OligoAnalyzer/) .

On the other hand, the hairpin structures of the rplK promoter sequences could have a detrimental influence on the transcription event by preventing RNA polymerase binding to specific recognition sequences in the promoter. However, this possibility also seems unlikely, at least when *E.coli* is concerned, because the single-stranded binding protein SSB is normally employed during replication and transcription to prevent the formation of any intramolecular secondary structures.

Regardless of the effect of the straphyloccocal promoter on CHMO expression in *E.coli*, the results for the NADPH oxidation assays showed that the vector pQR1030 allowed for the expression of CHMO expression in *E.coli* despite the presence of a second staphylococcal promoter downstream from pTac. In

addition, the results also showed an improvement in CHMO expression using pQR1030 when compared to the previously constructed expression vector pQR239.

4.7- Cloning of pQR1030 into *S. carnosus*

The vector construction strategy for pQR1030 had relied upon *E.coli* as the cloning strain. Consequently, after confirming that the vector granted expression of CHMO in the latter, the next step was to transfer pQR1030 into *S. carnosus*.

The transformation protocols described in previous studies for DNA cloning into the Gram-positive bacteria were primarily based on protoplast transformation (Gatermann & Marre 1989) and electroporation (Kraemer & Iandolo 1990). More recently, a protocol had been optimized for the development of electrocompetent *S.carnosus* (Löfblom et al. 2007), which was subsequently been used to clone pQR1030 into the staphylococcal strain (see chapter 2, section 2.2.5). Unfortunately, attempts to clone pQR1030 using this protocol were unsuccessful, and required further trouble-shooting.

Trouble-shooting the electroporation process was done by applying a design of experiments (DoE) approach, in which a large population of process variables was screened for the statistical importance of each variable on the overall process. Seven major variables were chosen from a list of possible factors affecting the electroporation protocol. These variables were growth media, starting cellular OD, electroporation medium, washing buffer, electroporation voltage, heat treatment, and incubation time after electroporation.

Previous studies on the electroporation process of several different bacterial strains had highlighted the importance of the cell density before the washing stages, with most protocols suggesting an optimal cell density after incubation of 0.5 OD (Gehl 2003). Alternately, in the optimized protocol developed by Löfblom *et al*. (2007), incubation time was identified as a crucial parameter, while variation in cell density did not have a big effect except below a certain level. Similarly,

incubation time after electroporation was highlighted as a crucial factor, since enough time had to be allowed for the cells to replicate the transformed plasmid.

By contrast, no studies on electroporation of *S. carnosus* had focused on the composition of the cell growth media, with different protocol using different types of media for culture incubation. B2 medium had been routinely used throughout this project to culture *S. carnosus*, but this media contained high concentrations of salt that could affect the translocation of negative-charged DNA molecules across the cell membrane.

There was also no consensus in the literature about the buffer used for the washing steps or the buffer in which the cells were resuspended for electroporation. Some protocols used distilled water, which increases the turgidity of cells and expands the cell membrane, while others used hypertonic solutions containing magnesium and sucrose that could be subsequently used as the electroporation buffer. The difference in approaches to the washing steps seemed to stem from different concerns that focused either on the survivability of the cells or on the efficiency of washing steps.

The two last variables selected for DoE analysis, electroporation voltage and heat-treatment of cells prior to electroporation, had been previously highlighted as crucial for the success of the process. Heat treatment in particular was introduced to the electroporation protocol by Löfblom et *al*. (2007) as a means to reverse the inactive metabolic process that could be responsible for rejection of foreign DNA. Alternately, voltage was identified as a important factor for the migration of plasmid DNA across the thick peptidoglycan layer of the staphylococcal bacterium.

A DoE two-level factorial design was conducted by running the seven variables concurrently at two different operational levels: a low (-) level below or corresponding to the optimal operation conditions of the process; and a high (+) level that was set at the higher extreme of the window of optimal operation conditions. Thus, the seven variables were combined into 32 experimental runs with alternative levels for each of the variables (Appendix 3). Results were measured as number of surviving colonies on plates containing the antibiotic chloramphenicol after transformation with pQR1030. Plasmid concentration, incubation temperature, shaking speed, and sample volumes were maintained

constant throughout the run. This type of factorial design is a very powerful tool for the optimization of an experimental process because it provides statistically significant data relative to the influence of individual variables to the overall performance of the process, while also providing a statistical map of the functional relationships between the different tested variables.

Figure 45 shows a graphical representation of the results from the DoE analysis in the form of a Pareto chart. The y-axis of this chart corresponds to the effect values of the individual variables or combinations of two variables, which in turn relates to their statistical significance to the performance of the experimental proccess. Factors with high effect values are considered to be more significant for the process, and changes in these generate large fluctuation in the overall performance. The results from the analysis highlighted 3 variables that were significant for the electroporation protocol: initial OD, electroporation voltage, and incubation time after electroporation.

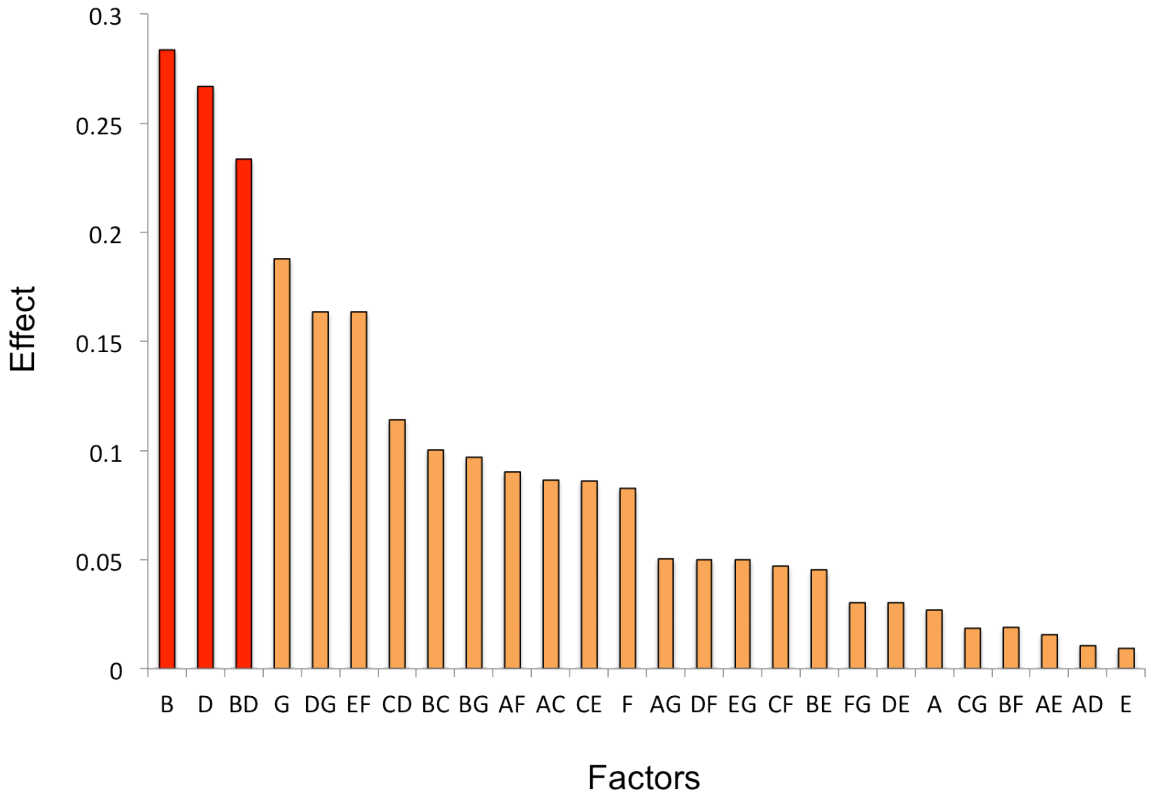


Figure 45- Pareto chart showing the effect levels of different factors individually and combinations of two factors to the efficiency of *S. carnosus* electroporation. Effect values are expressed as the fraction between factors and the sum of the effect values of the whole factor population. The 7 factors are designated as letters, from A to G: A- Washing buffer; B- Electroporation voltage; C- Growth media; D- Starting cell density (OD); E- Heat treatment; F- Electroporation medium; G- Incubation time. The factors B and D, together with the interaction between these two factors, generated the highest effect values and are highlighted in red.

Increasing the starting cellular OD from 0.1 to 0.5 and incubating the cells for 2 hours after electroporation had a positive effect on the transformation efficiencies. The importance of these variables can be readily explained. By starting with a higher OD, the cell concentration of the electroporation samples was increased, which in turn increases the probability of successful transformation to occur. Similarly, the longer incubation time after electroporation allowed for more replication of the plasmid to occur inside of cells, thereby increasing the survivability of cells in antibiotic selective plates.

On the other hand, electroporation voltage affected the system inversely, as raising the voltage resulted in decrease in the efficiency of electroporation. The optimized protocol devised by Löfblom *et al.* (2007) had identified the voltage of 2.1

kv/cm as the optimal for electroporation, and consequently increasing the voltage would compromise the survivability of the cells. The results from the DoE analysis corroborated this assumption.

On the other hand, the DoE analysis did not identify heat-treatment, wash buffer, or electroporation buffer as factors that significantly impacted the electroporation process.

The results from the DoE analysis led for a re-design of the electroporation protocol that achieved reproducible transformation of *S. carnosus* with the vector pQR1030. However, the transformation efficiencies were still very low and required high plasmid DNA concentrations.

An alternative method of transformation based on sonication (L. Lin et al. 2010) was also tested. This method was faster than the electroporation protocol, as it did not rely on lengthy incubation periods and wash steps prior to transformation. As Gram-positive bacteria, *S. carnosus* has a structurally robust cellular wall that is resilient to physical shear forces. Taking advantage of this natural resilience, the sonoporation protocol employed ultrasound to create pores on the membrane of cells without compromising the general structure of the bacteria.

Using this method, cells could be directly transformed with plasmid vector after overnight growth in B2 medium. The resulting bacterial culture was aliquoted into 500 µl samples and mixed with plasmid DNA. Samples were then sonicated at a high frequency for 20 seconds, and subsequently re-grown in 5 ml of B2 medium for 4.5 hours.

Transformation efficiencies using this method were comparable to the transformation rates of electroporation. Since it did not require complex and time-consuming pre-treatment steps, this sonication-based method became the preferred protocol for vector transformation into *S. carnosus*. A detailed description of the protocol can be found in the materials and methods chapter (see chapter 2, section 2.2.6).

4.8- pQR1030-mediated Expression of CHMO in *S. carnosus*

S. carnosus samples transformed with pQR1030 were initially tested for CHMO expression by using the NADPH oxidation assays previously described for *E.coli*. Due to the fact that *S. carnosus* was more resilient to physical stress than the Gram-negative counterpart, a higher pressure and increased number of passes needed to be applied in the French Press homogenization proccess, in addition to a chemical lysis step with lysostaphin at 37 ^{0}C.

Figure 46A shows the results from these experiments. No activity could be detected in *S. carnosus* samples grown for 8 hours in medium containing the antibiotic chloramphenicol. This suggested that while pQR1030 was being actively replicated, CHMO was either not being expressed inside the cells or that it was not active. As this protein is not native for Gram-positive bacteria, there could be the possibility that it was not stable inside of *S.carnosus* and would subsequently aggregate into inclusion bodies or be degraded due to miss-folding. Aggregation could also be a result of over-expression, and since the staphylococcal *rplK* promoter was located upstream of a ribosomal gene, there was the possibility that this promoter was promoting strong constitutive expression of the biocatalyst.

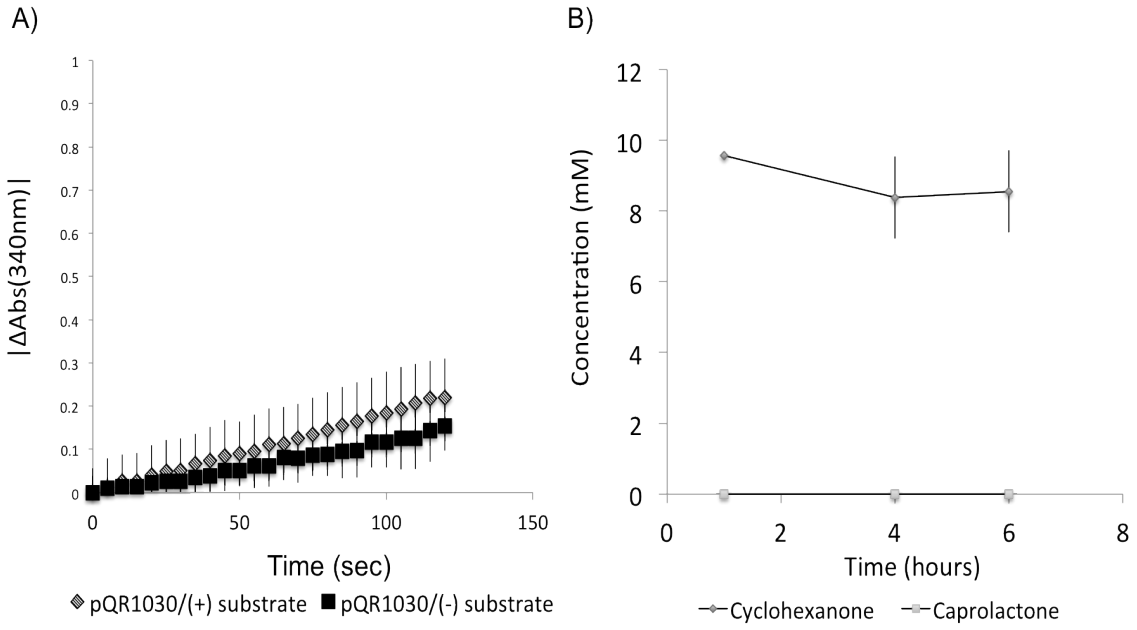


Figure 15- Biocatalysis assays with *S. carnosus* TM300 containing the vector construct pQR1030. A) NADPH oxidation assay on cell lysates. B) GC analysis of cyclohexanone and caprolactone levels in biocatalytic reactions conducted on resting cells. Error bars refer to the standard deviation of the experimental set-up, calculated from triplicates of the same assay.

To test these hypotheses, the supernatant and pellet fractions from the lysed cell cultures were loaded unto an SDS-PAGE gel. The results from this experiment showed no difference in the protein profiles between the samples containing the pQR1030 vector and the control *S. carnosus* samples that did not contain any plasmid. Over-expression of CHMO would result in a intense band around the 60 KB size range after staining, but not difference in intensity of the bands around that size were detected on the samples containing the expression plasmid.

Finally, to test if the pQR1030-containing cultures could perform whole-cell biocatalysis of the substrate cyclohexanone, 50 ml batches of cultures grown overnight were resuspended in 5 ml of Tris-HCl buffer (pH 7.4) containing 10 g/L of glycerol and 10 mM of cyclohexanone and subsequently incubated for 6 hours. Production of the product caprolactone was monitored through the incubation period through GC analysis (figure 46B). The results showed that there was no production of caprolactone during the biocatalytic reaction, despite the slight

decrease in the concentration of cyclohexanone after the incubation period, which was attributed to the evaporation of the volatile compound.

Together, the results from the GC analysis, NADPH oxidation assays, and SDS-PAGE suggested that either the CHMO enzyme was not being actively expressed inside *S. carnosus* , despite the presence of the putative staphylococcal rplK promoter upstream from the gene. It was subsequently concluded that this staphylococcal promoter had not been correctly selected as a strong genomic promoter for heterologous protein expression.

4.9- Introduction of Reporter Gene into pQR1030

The inability of pQR1030 to express the CHMO protein in *S. carnosus* highlighted the inherent risk in the vector design strategy employed for the construction of the vector. Choosing not to rely on staphylococcal promoters that were already characterized in the literature meant that expression of the CHMO gene could not be guaranteed through the methodology used for promoter selection. In fact, the *in-silico* tools used to screen the staphylococcal genome for strong promoters were predictive in nature and did not have a solid empirical database to support them. Thus the probability of choosing non-functional sequences or sequences that revolve around complex regulatory systems was very high. Fortunately, the modular structure of the vector was designed to account for this risk through the inclusion of two distinct restriction sites at the end of the staphylococcal promoter that permitted its replacement.

Therefore, the most accessible solution to the non-expression of CHMO in *S. carnosus* would be to select and test new set of genomic promoters that would replace the rplK promoter as active variants. However, using CHMO as a reporter gene for the screening of strong promoters was not ideal. As pointed out previously, most direct ways to measure CHMO activity required some amount of downstream processing that involved time and work-intensive lysis steps. A more rapid an immediate way to screen for active promoter sequences in *S. carnosus*

would rely on a reporter gene that gave a phenotypical feedback, such as GFP (green fluorescent protein) or an antibiotic resistance gene. The feedback would provide a quick and direct method of measuring the strength of a promoter without the need for downstream-processing protocols.

A variant of pQR1030 was thus constructed, in which the CHMO gene was replaced by the gentamicin resistance gene, aminoglycoside N(3')-acetyltransferase I. This gene was chosen because it was readily available, and although not being a visual reporter like GFP, it granted a phenotypical change to the bacteria that can be directly measured by using gentamicin as a selective antibiotic in the growth medium.

The cloning strategy for this reporter variant, designated pQR1031, was slightly different from the method used to construct pQR1030 (Figure 47). The gentamicin resistance gene (GenR) was amplified from the vector pR26 using primers with *Xho*I and *Sph*I site overhangs that allowed for cloning into the MCS of pQR1030, but the PCR product was subsequently cloned into a commercial TOPO vector (Invitrogen) prior to insertion into the expression vector, instead of directly cloning the amplified fragment into pQR1030. This extra step was done for two reasons: firstly, some restriction endonucleases require extra base-pairs upstream and downstream of the restriction sites to cut the DNA efficiently, and the primers used for the PCR amplification were not designed with this in mind; secondly, the PCR amplification of GenR was not very efficient and resulted in low concentrations of the amplified product. Thus, using the blunt-end high-copy number TOPO commercial plasmid as a carrier for GenR ensured that the restriction with *Xho*I and *Sph*I worked, and also maximised the concentration of GenR by using *E. coli* cultures as the source of the gene.

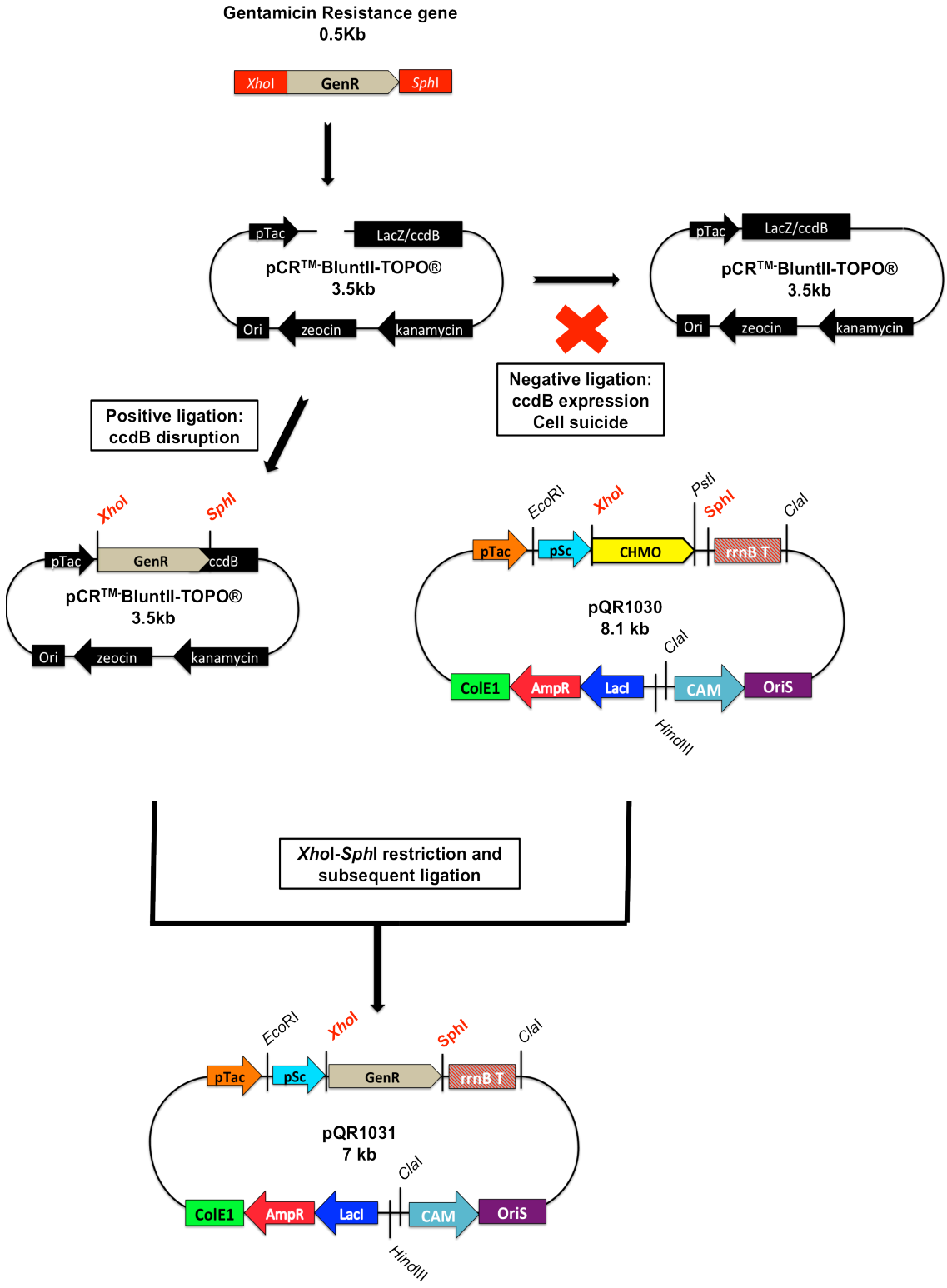


Figure 47- Ligation strategy for cloning of the gentamicin resistance gene (GenR) into pQR1030. The first step involved cloning the PCR-amplified gene into the commercial pCR™-BluntII-TOPO® plasmid, which allows for blunt-end cloning. Upon ligation, the inserted gene disrupts the expression of the suicidal gene ccdB, thus allowing the survival of cells transformed with the positive ligation. The GenR gene is subsequently cut with *Xho*I and *Sph*I and ligated to the pQR1030 construct, which had been cut with the same enzymes in order to remove the CHMO gene.

In turn, the ligation of the gene into the expression vector was aided by the fact that the restriction of pQR1030 with *Xho*I and *Sph*I was easily accessed, as it generated two distinctly sized bands corresponding to the backbone of pQR1030 and the smaller CHMO gene. After endonuclease restriction, restricted samples were loaded unto an agarose gel and purified by gel extraction. Ligation reactions of the pQR1030 backbone with GenR were performed with T4 ligase (NEB) and incubated for 1 hour at 37 ^{0}C. The ligation samples were subsequently transformed into chemically competent *E. coli* TOP10 cultures, which were selected on gentamicin containing plates. In theory, only successful insertion of GenR into pQR1030 would grant the gentamicin resistance phenotype through pTac-mediated expression of GenR.

Figure 48 shows the screening results from plates of *E.coli* colonies transformed with the ligation product. Cutting the ligation product pQR1031 with XhoI and *Sph*I generated two bands of different sizes that corresponded to GenR (533 bp), and the plasmid backbone (6.5 kb). Samples containing the correct restriction map were also sequenced to confirm that GenR was ligated in frame.

Once the ligation was confirmed, the vector was transformed into *S. carnosus* and selected on chloramphenicol plates. Since the vector pQR1030 did not express the CHMO gene inside the Gram-positive bacteria, it was not expected that colonies transformed with pQR1031 would be gentamicin resistant.

To test if the gentamicin resistance gene could work as a reporter gene to measure promoter strength, both pQR1031- containing *E.coli* and *S.carnosus* cultures grown overnight were plated on agar plates with increasing concentrations of gentamicin. Results were expressed as number of surviving colonies after overnight incubation (Figure 49), and demonstrated that this gentamicin resistance assay was sensitive enough to detect variations of protein expression over small gentamicin concentration gradients.

The colony survival rates for pQR1031-containing *E.coli* mirrored previous expression assays with pQR1030, in which the tight control of the pTac promoter by *Lac*I resulted in very low levels of CHMO expression under non-induced conditions.

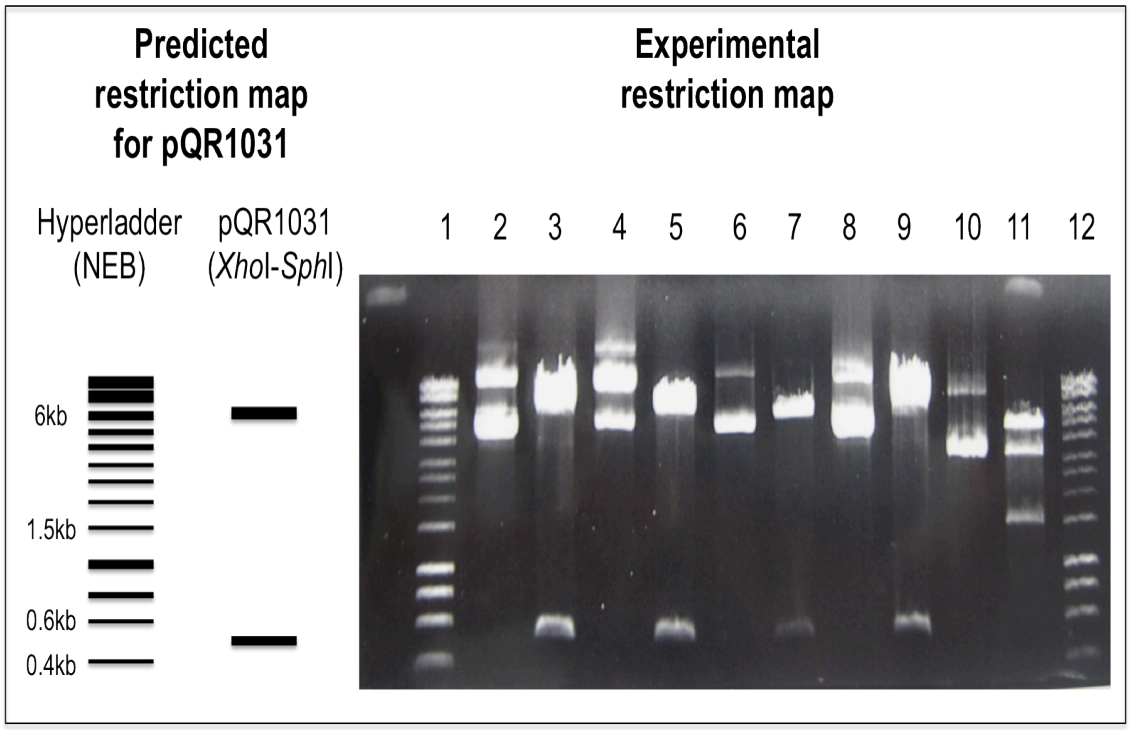


Figure 48- Agarose gel (0.7% w/v) showing the results from a restriction digest of plasmid samples from *E.coli* TOP10 transformed with the ligation products. The predicted restriction map of the positive construct pQR1031 is also displayed. Positive ligations would be expected to genereate 2 bands when cut with the *Xho*I- *Sph*I restriction mix, the lower of which corresponds to the GenR gene.

On the other hand, the results obtained from the *S. carnosus* samples were unexpected. Indeed, *S. carnosus* strains containing the pQR1031 vector were not only able to express gentamicin resistance, but they were also able to survive higher gentamicin concentrations than their *E. coli* counterparts. Contrary to the studies conducted on pQR1030, these results suggested that the staphylococcal rplK promoter was active inside *S. carnosus*.

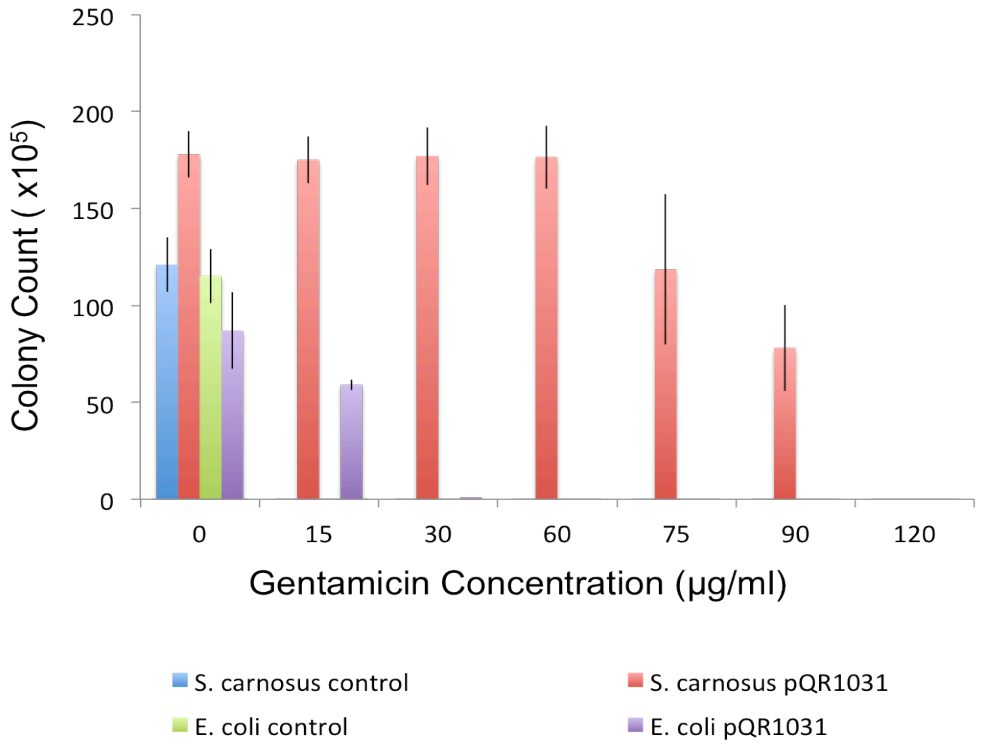


Figure 49- Gentamicin resistance assay with *E. coli* TOP10 and *S. carnosus* TM300 containing the pQR1031 construct, as well as with strains not containing any antibiotic resistance. The resistance to the antibiotic is expressed in number of colonies in plates containing increasing amounts of gentamicin. Error bars refer to the standard deviation of the experimental setup, calculated from triplicates of the same assay.

More relevant to the project was the fact that replacing CHMO with the gentamicin resistance gene allowed for the detection of expression from the *rplK* promoter in *S. carnosus*. While this promoter could not be considered a high expression constitutive promoter, the difference in levels of gentamicin resistance between the vector-containing *S. carnosus* strain and un-induced *E.coli* strain suggested that CHMO activity should have been detected in lysates of pQR1030-containng *S. carnosus*. The inability to detect any CHMO activity from staphylococcal lysates favoured the hypothesis that CHMO was not being stably expressed inside *S. carnosus.*

Regardless, the gentamicin resistance assays validated the vector pQR1031 as a reporter system that could be used for the screening of libraries of staphylococcal genomic promoters.

4.10- Designing Efficient *S. carnosus* Promoters

The newly constructed reporter vector pQR1031 provided the required tools to rapidly and accurately screen large libraries of staphylococcal genomic promoters. In order to create these libraries, the promoter selection strategy needed to be revised to account for some of its limitations.

The initial promoter selection strategy relied upon the use of sequences upstream from the predicted HEG that could potentially contain strong promoters. These sequences were not screened for specific promoter elements but instead used entirely as a single expression unit. This was the case of the *rplK* promoter, which was actually the upstream sequence from an putative operon containing the ribosomal 30S subunit L-11. The motivation to use entire sequences upstream from HEGs was twofold: on one hand it was difficult to pin-point specific promoter elements within the sequences because these elements varied greatly from consensus bacterial elements in Gram-negative bacteria; on the other hand, these sequences upstream of HEGs could contain multiple promoters that would add up to the overall transcription activity of the sequence.

However, there was also a risk associated with the inadvertent selection of sequences containing regulatory regions of unknown function that could interfere with promoter activity. Since the transcription regulation of S*. carnosus* has not been well characterized, the probability of selecting for un-identifiable regulatory sequences was high.

To mitigate this risk, the promoter strategy was re-designed by applying analytical filters to the promoter selection analysis done previously. Sequences that had been highlighted previously as potential promoters were actively scanned for promoter consensus motifs using promoter recognition software (BDGP- http://www.fruitfly.org/seq_tools/promoter.html, BPROM- http://linux1.softberry.com/). In particular, the software searched for regions within sequences that shared similarities with the consensus -35 and -10 elements of a bacterial promoter that operated under σ^{70} RNA polymerase transcription. In

addition, it also accounted for the optimal distance between these elements, whereupon only segments of sequences that contained -35 or -10-like elements at 17 base pair distance from each other would be considered as potential promoters.

In reality, there is a broad range of variation to promoter regions outside the Gram-negative based consensus used by the software. There have been some studies in the literature pointing out the fact that promoter regions in Gram-positive bacteria are not as tightly conserved and elements other than the -35 and -10 recognition sites within the promoter might be more crucial and conserved within the Gram-positive promoter (Voskuil 1998; Schofield et al. 2003). For instance, an A-rich region upstream of the -35 element and a smaller TXTG motif just upstream of the -10 element have been reported to have a function in promoter recognition of gram-positive expression systems, and also have been related to the inability of promoters with tightly conserved -35 and -10 elements like pTac to express in Gram-positive bacteria. One could thus argue that using software that is based on a Gram-negative promoter consensus is inadequate to identify promoter regions within gram-positive genomes.

However incomplete this *in-silico* analysis might be, it was still useful as a tool to highlight windows within *S. carnosus* genomic sequences that might contain promoter activity, which could then be complemented with the knowledge of the other motifs that are also present in Gram-positive promoters.

Figure 50 shows the revised promoter selection strategy. In the first steps of the strategy, the library of HEGs identified during the first run of promoter selection was re-used to identify sequences upstream that could be used as potential promotes. These sequences were subsequently filtered through open-source software that screened specifically for smaller fragments containing promoter elements that fitted several sequence-specific criteria, such as the distance between promoter elements and the presence of secondary motifs like AT-rich sequences around the promoter elements. These smaller promoter-like fragments were then re-designed with additional genetic elements to fit a more conventional promoter structure. In many cases the addition of a ribosomal binding site (RBS) and restriction enzymes was required.

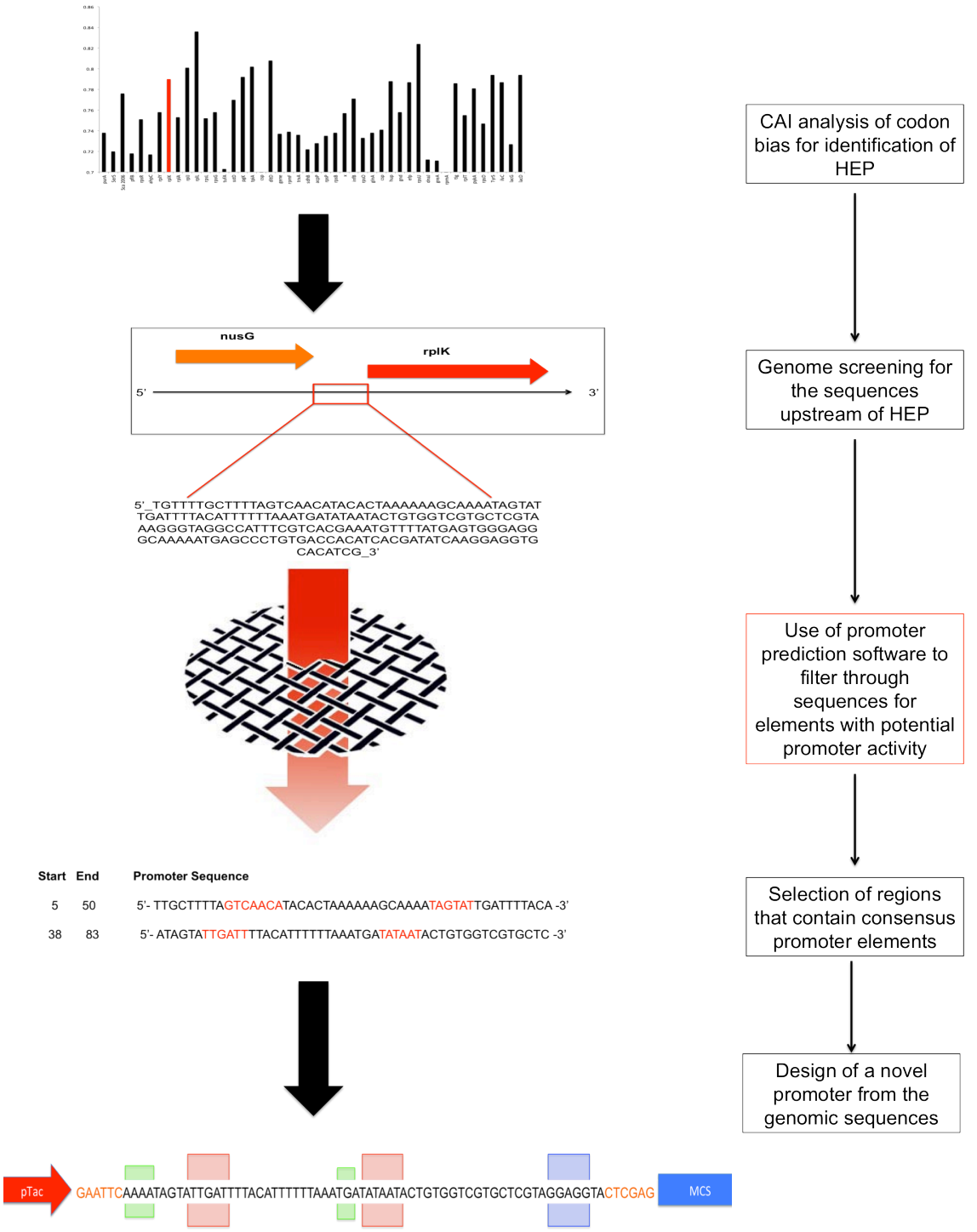


Figure 50- Strategy for screening the *S. carnosus* genome for putative promoters, in which an intermediate step is applied to filter sequences upstream of putative highly expressed proteins (HEPs) for specific promoter elements. The shortes sequences containing these elements are subsequently designed as 80 to 90 bp promoters by including a RBS downstream and restriction sites to allow for the cloning into the synthetic promoter region of the shuttle vector.

Using this approach to promoter selection, two putative promoters were designed with sequences taken from areas of genomic DNA upstream of the 50S ribosomal protein L11 (*rplK*) and the 50S ribosomal protein L10 (*rplJ*). The choice to continue working with the sequence upstream of the protein *rplK* , even after the poor performance of the vector pQR1030, was based upon the idea that the revised *in-silico* screening strategy should be able to identify shorter sequences within the rplK upstream region with promoter activity. Therefore, re-designing the rplK promoter used in pQR1030 was a way to validate the novel promoter selection strategy. In addition, comparison between the activity of the rplK promoter used in pQR1030 and the smaller re-designed iteration would be valuable in accessing the elements within the staphylococcal region that inhibited pQR1030-mediated CHMO expression.

The sequence immediately upstream of the protein *rplJ* had been previously highlighted as an area of potential strong promoter activity during the first run of promoter selection. The high CAI values calculated for *rplJ* and another ribosomal protein, *rplL*, in close proximity (Figure 51) indicated that the area just upstream of both proteins could arbor promoters that catered for two highly expressed proteins. In fact, the close proximity between the two genes and the same direction of the open reading frames (ORF) suggested that these proteins might in fact operate as an operon unit. If this is the case, there is a possibility that a strong promoter was inducing the simultaneous transcription of both ribosomal proteins.

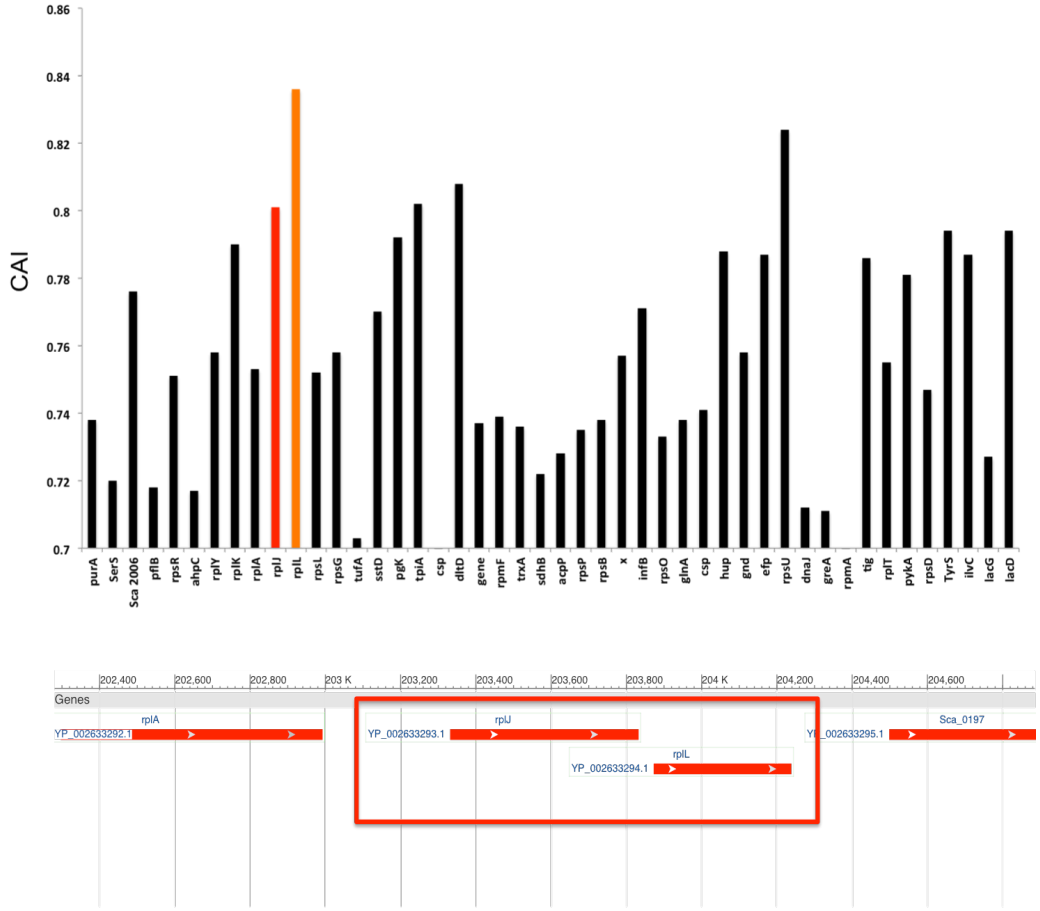


Figure 51- CAI levels of a set of proteins from *S. carnosus* , together with the graphical representation of the ribosomal proteins rplJ and rplL as organised in the genome of *S. carnosus* (highligted by the red outline). These genes are highlighted in the barchart as a red and yellow bar, respectively.

Both of the re-designed promoters were made up of a main fragment selected from the genome of *S. carnosus*, and a RBS, which was also lifted from sequences immediately upstream from the starting codons of the rplK and rplJ genes. This site was designed 7-8 bps upstream from the protein starting codon to allow for efficient translation. *Eco*RI and *Xho*I restriction sites were also added to the ends of the promoters to allow for the replacement of the rplK promoter in pQR1031 (Figure 52).

In addition to the -35 and -10 promoter recognitions sites, sequences upstream of HEGs were also actively screened for nucleotide motifs around these sites that had been previously identified as relevant for Gram-positive promoters. These include a poly-A sequence located 42 bps upstream from the start of

transcription (-42), and a dinucleotide TG motif located around at 16 bps upstream from the start of transcription.

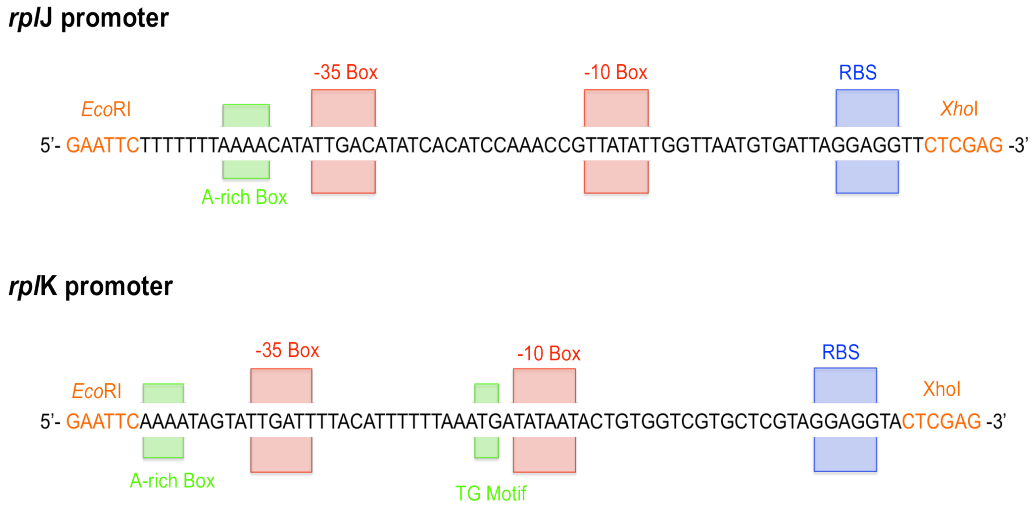


Figure 52- Sequences of the designed *rplJ* and *rplK* promoters. The -35 and -10 promoter elements are contained within red boxes, while the RBS sequences, which were taken from upstream of the respective ribosomal genes, are highlighted by the blue boxes. The green boxes contained short-sequence elements that had been previously identified as important elements in gram-positive promoters.

The promoters were commercially synthetized as single-stranded short oligos that self-annealed upon incubation at 55 ^{0}C. To ensure efficient restriction of the ends of promoters with the *Xho*I and *Eco*RI endonucleases, the promoters were subsequently self-ligated using a blunt-end ligation protocol. This step resulted in continuous DNA strands of repeating promoter units, separated by restriction sites that had enough extra base pairs at the ends to allow for restriction with the corresponding endonucleases.

After restriction, the re-designed promoters were ligated into the *Eco*RI/*Xho*I restricted backbone of pQR1031, which was gel extracted prior to ligation. Ligation was performed with T4 ligase, with ligation reactions being incubated for 1 hour at 37 ^{0}C. Ligation products were transformed into chemically competent *E. coli*, which were subsequently selected for positive ligations in ampicillin selective plates. Positive ligations were confirmed from sequencing of the area around the staphylococcal promoter region of the vector (Appendix 1), as the insertion of the

smaller promoters into pQR1031 was impossible to access through restriction maps due to their size.

The new constructs pQR1032 and pQR1033, containing the re-designed rplK and rplJ staphylococcal promoters respectively, were subsequently transferred into *S. carnosus* and tested for promoter activity.

4.11- The Expression Profiles of pQR1032 and pQR1033 in *E. coli* and *S. carnosus*

After successful transformation into *S.carnosus* TM300 and *E.coli* TOP10, the new promoter designs were tested for expression of GenR. In a similar assay to the antibiotic resistance assay developed for pQR1031, individual colonies containing pQR1032 and pQR1033 were incubated in 2 ml square 96-well plates containing liquid media with increasing concentrations of the antibiotic gentamicin for 24 hours at 37 ^{0}C and 900 rpm. The levels of bacterial resistance to the antibiotic were measured as a function of OD at 600 nm.

As shown in Figure 53, this assay produced very different results from the tests done with pQR1031. Accordingly, *E. coli* strains containing the new constructs were able to survive in media containing higher concentrations of gentamicin than the pQR1031-containing strain. Since the experiments were conducted in the absence of IPTG, and under these conditions pTac-mediated expression is greatly inhibited by *Lac*I, it could be concluded that levels of survivability observed were a product from the activity of the re-designed staphylococcal promoters.

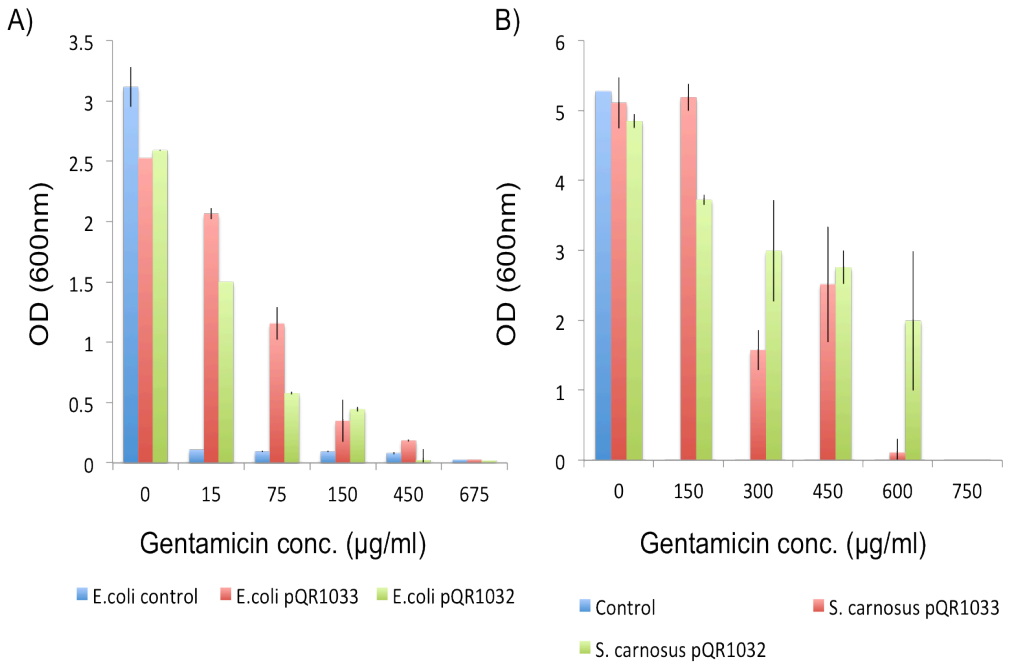


Figure 53- Gentamicin resistance assays performed on *E. coli* TOP10 (A) and *S. carnosus* TM300 (B) containing the pQR1032 and pQR1032 constructs. Values on the y-axis are expressed as optical density of cultures grown overnight at 600 nm. Error bars refer to the standard deviation of the experimental setup, calculated from triplicates of the same assay.

The activity of the *rplK* and *rplJ* promoters in *E.coli* was unexpected, since the core of these promoters had been lifted from a Gram-positive genome, which is a genetically distinct system from its Gram-negative counterparts. However, the design of the new promoters was informed by a promoter consensus structure that had been based primarily on similarities between Gram-negative promoters, and therefore the sequences lifted from the *S. carnosus* genome might have promoter recognition sites that are easily recognized by the *E. coli* transcription apparatus. These results also corroborated previous studies in which *E.coli* was shown to have enough plasticity to work with Gram-positive expression systems.

The constructs pQR1032 and pQR1033 also resulted in an almost 10-fold increase in *S. carnosus* resistance to gentamicin compared to the pQR1031 vector. These results not only validated the revised promoter selection strategy as an efficient way of constructing libraries of strong constitutive promoters based on the genome of the host cell line, but also corroborated the previous hypothesis that the longer staphylococcal promoter sequence in pQR1031 contained genetic elements that were detrimental to protein expression.

Comparison between the levels of survivability of pQR1032 and pQR1033-containing *S. carnosus* indicated that the re-designed rplK promoter was more active than the rplJ counterpart. This disparity might be explained by differences between the sequences of the promoters. The rplK sequence contains the complete -10 TATAAT box as well as TG motif at position -16 that has been previously linked with successful transcription in the Gram-positive *Bacillus subtilis* Inversely, the *rpl*J promoter contains an intact -35 box (TTGACA), while most other promoter elements in the sequence are not as well conserved. This suggests that the -10 and -16 recognition sites are more important for promoter strength in *S. carnosus* than other elements further upstream.

On the other hand, a comparison between the survivability levels of *E. coli* and *S. carnosus* strains highlights the advantage of using promoter systems that are native to the host organism. Despite being active in *E. coli*, the pQR1032 and pQR1033 exhibited higher GenR expression level in *S. carnosus*, from which the rplK and rplJ promoters are native to.

In the end, by using a combination of *in-silico* tools that focused on highly expressed proteins and on elements that form the structure of a bacterial promoter, a library of short genomic promoters was created that could be used for strong heterologous protein expression in both *E. coli* and *S. carnosus*. This ability to screen the host genome for strong promoter elements has immense potential as a tool for the creation of promoter libraries that fit a variety of protein expression requirement, but unfortunately that potential lies outside of the initial aims of the project, i.e. to provide an efficient expression system for CHMO in *S.carnosus*.

Consequently, the better of the expression constructs, pQR1032, was subsequently used as the expression vector for the CHMO gene.

4.12- Concluding Remarks

The construction of a strong expression system for CHMO expression in *S. carnosus* was more time-consuming and problematic than first expected. Part of the problem stemmed from the quality of the genetic materials that composed the

system. For instance, the first step in the ligation process involved the ligation of a synthetic promoter cassette to the vector pTTQ18, and since the synthetic component was provided commercially as a pure product ligation was successful on the first attempt. On the other hand, subsequent ligation steps required the development of several additional protocols to ensure that the genetic material being handled was provided in sufficient amounts or with sufficient purity for ligation to occur. In particular the handling of genetic material in *S. carnosus* required the creation of efficient processes that would facilitate plasmid extraction and transformation. While the development of these processes was time-consuming and required extensive trouble-shooting, they allowed for an extensive characterization of the operational limits of *S. carnosus* as a cloning system, and provided us with tools that could be reliably used. For example, establishing a method for the transformation of *S. carnosus* via sonication proved to be an essential for the subsequent transfer of the different expression vectors into the Gram-positive bacteria.

Another problem with the vector construction strategy was the decision not to rely on promoters that already been proven to work in *S. carnosus*, like the lipase promoter from the vector pNW21 or the inducible xylose promoter used routinely in many other expression systems (Wieland et al. 1995). The choice to screen for novel promoters in the genome of *S. carnosus* originated for a resolve not to be constrained by the limitations of these conventional systems, but as a consequence increased the risk of ending up with promoters that did not fit the necessary requirements, as was the case with the long *rpl*K promoter.

In-silico tools that allow for the screening of particular elements in the genome have an important role in mitigating this risk. The fact that we needed to revise our initial screening approach is a testament to limitations of *in-silico* models as predictive tools when trying to account for complexities of a biological system in the absence of experimental evidence. However, by coordinating different *in-silico* approaches we were able to increase our predictive ability, and as a result succeeded in discovering sequences in the *S. carnosus* genome that could work as strong constitutive promoters.

In the end we fulfilled the goal of creating an efficient expression vector, while also creating in the process several derivatives of this vector that could be

used as rapid screening tools for the mining of promoters from *S. carnosus* and other related bacteria. The next step in the project was to re-clone the CHMO gene into the expression vector pQR1032 and to characterize its expression in *S. carnosus*.

Chapter 5: Results- Study of *S. carnosus* as a Viable Host for CHMO Biocatalysis

5.1- Expression of CHMO in *E.coli* TOP10 using the Expression Vector pQR1034.

The cloning of the CHMO gene into the vector construct pQR1032 was a straight-forward process that involved the replacement of the GenR gene with the amplified CHMO gene via a *Xho*I/*Sph*I restriction and subsequent ligation. A similar cloning strategy to the construction of the reporter vector pQR1031 was employed for this ligation (see chapter 4, section 4.9). The CHMO was amplified from the vector pQR239 with primers containing the necessary restriction sites at the ends, and cloned into a pCR^{TM}-BluntII-TOPO® plasmid (Invitrogen) prior to restriction in order to maximize restriction efficiencies. After the *Xho*I/*Sph*I restriction, both the CHMO gene and the backbone of pQR1032 were gel extracted, and subsequently ligated using T4 ligase (NEB). Ligation reactions were incubated for 1 hour at 37 ^{0}C. The new CHMO expression construct, pQR1034 (Figure 54), was transformed into *E. coli* TOP 10 and confirmed via restriction maps and sequencing (Appendix 1).

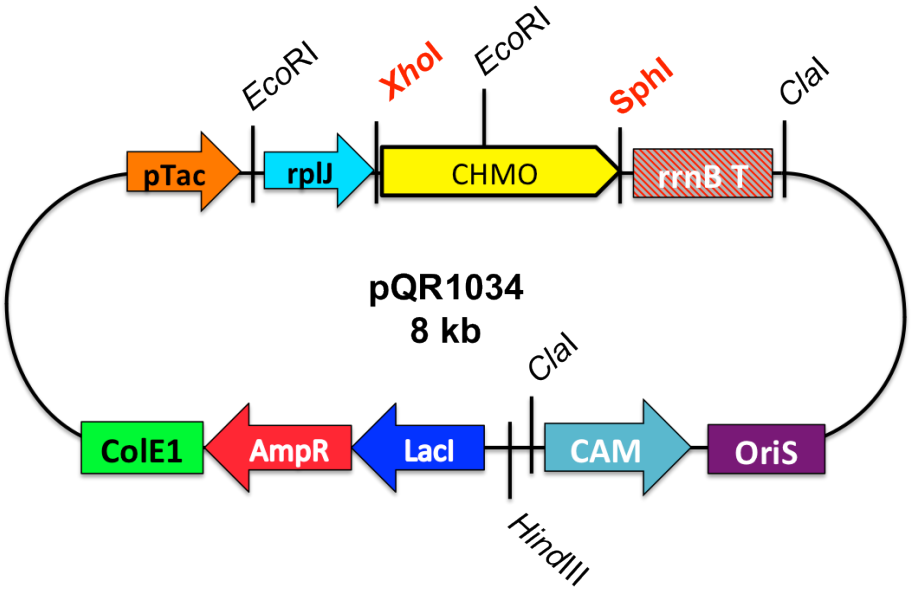


Figure 54- Diagram of the pQR1034 expression vector. The genes and promoters in the diagram do not represent the actual size of the genetic elements. The CHMO gene was cloned into the plasmid pQR1032 via a *Xho*I-*Sph*I restriction and subsequent ligation (sites highlighted in red).

The efficiency of the novel expression vector was subsequently tested on *E. coli* TOP10 to access if the promoter change had indeed resulted in optimized expression of CHMO . From the gentamicin resistance assays done previously (see chapter 4, section 4.10), it was expected for the re-designed rplK promoter to be constitutively active in *E. coli*, consequently adding to the levels of CHMO already expressed under the pTac promoter. The levels of CHMO expression were measured by performing NADPH oxidation assays on the lysates of vector-containing cultures. The detailed protocol for these assays can be found in the materials and methods chapter (see chapter 2, section 2.3.4). Figure 55 shows the results from NADPH oxidation assays done on E. *coli* TOP10 containing pQR1034, together with results obtained for pQR239 and pQR1030 constructs under the same experimental conditions (see chapter 4, section 4.6).

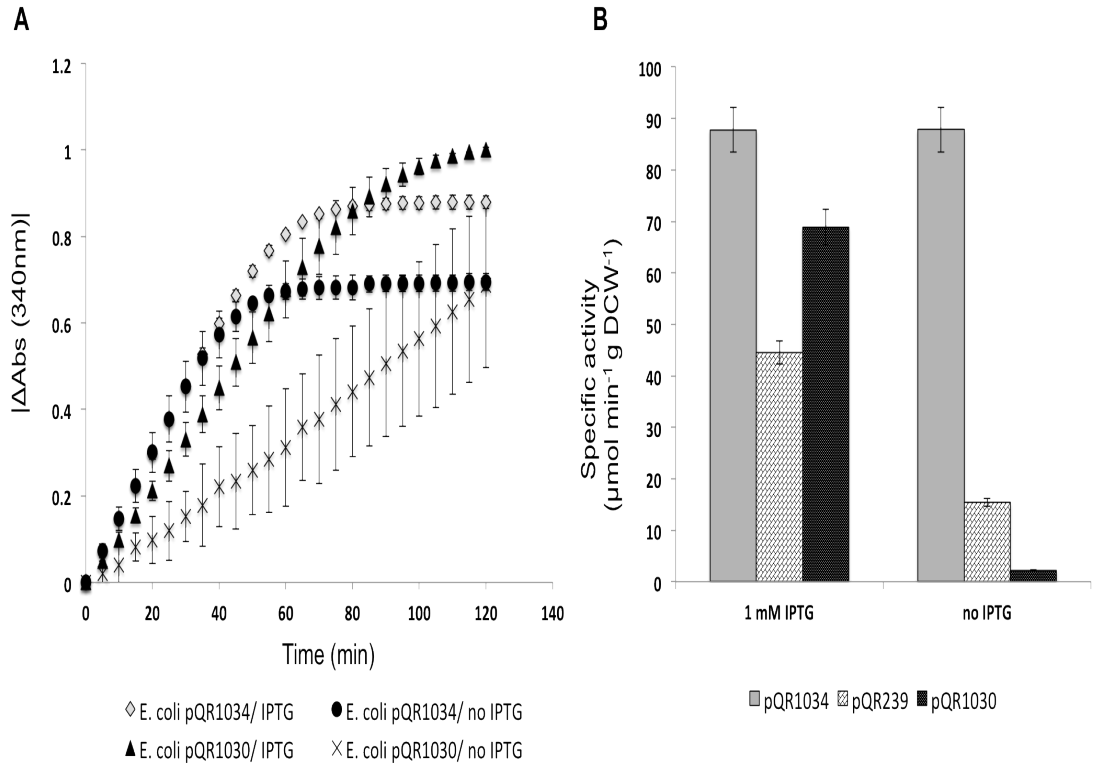


Figure 55- NADPH oxidation assays with *E. coli* TOP10 containing the vectors pQR239, pQR1030, and pQR1034. A) NADPH oxidation profiles of biocatalytic reactions performed with the lysates of the different expression strains. The y axis expresses the absorbance gradient between different time points of the reaction as an absolute number. In reality, the gradient is negative because of NADPH oxidation. B) Specific enzymatic activity of CHMO in the lysates, calculated as μmol of NADPH oxidized per minute per gram of dry cell weight (DCW). Error bars refer to the standard deviation of the experimental setup, calculated from triplicates of the same assay.

The activity profiles of the different constructs highlighted the effects of the different staphylococcal promoters on CHMO expression. As discussed previously, the CHMO activity profile of pQR1030-containing cultures was comparable to the control pQR239 strain under IPTG-induced conditions, while in the absence of inducer the *Lac*I repressor gene echibited a tight control on the activity of pTac in pQR1030 (see chapter 4, section 4.6).

By contrast, the pQR1034 construct resulted in a significant increase to the activity of CHMO in *E. coli* lysates, at around 88 μmol min^{-1} $gDCW^{-1}$ two-fold higher than pQR239. This reflected the results obtained with the gentamicin resistance constructs, and corroborated the previously drawn conclusion that the re-designed rplK vector expressed constitutively in *E. coli*. Furthermore, the addition of the

smaller staphylococcal promoter negated the tight expression control observed with pQR1030, as there is no difference in activity between pQR1034 samples in the presence or absence of IPTG.

The differences in activity between the pQR1034 and the pQR1030 lysates might account for the individual activities of both pTac and rplK promoters. Since the former is tightly controlled by the LacI repressor system, the CHMO activity observed in lysates of the pQR1034 construct might be conclusively attributed to the combined activity of the pTac and rplK promoter.

While CHMO has not been previously reported to be toxic to *E. coli* when overexpressed, there was still the possibility that the increased expression of CHMO using pQR1034 could be inhibiting to cell growth. To study this hypothesis, the growth kinetics of *E. coli* TOP10 containing the expression vector pQR1034 were compared to the same strain containing the plasmid construct pQR1030. Colonies containing the different vector constructs were grown overnight in 5ml of LB medium, and subsequently transferred to 250ml baffled shake flasks containing 50 ml working volume of LB medium with 100 µg/ml of ampicillin. The inoculated shake flasks were then incubated at 37 ^{0}C and 250 rpm for 8 hours.

Figure 56 shows the growth curves of *E. coli* TOP10 strains containing the CHMO expressing vector pQR1034 and the tightly controlled vector pQR1030. As it can be seen from the results, the maximum growth rates (μ_{max}) were comparable between the two strains, albeit the pQR1034 took longer to start exponential growth. Alternately the total biomass concentration was lower for the pQR1034 strain, at 0.73 g DCW when compared to 0.93 g DCW achieved by the pQR1030 counterpart. Together, these results suggested that the increased constitutive overexpression of CHMO induced by the staphylococcal rplK promoter negatively affected the growth of the *E. coli* host. This could be explained by a stronger metabolic burden exerted by the constitutive protein expression on the cell.

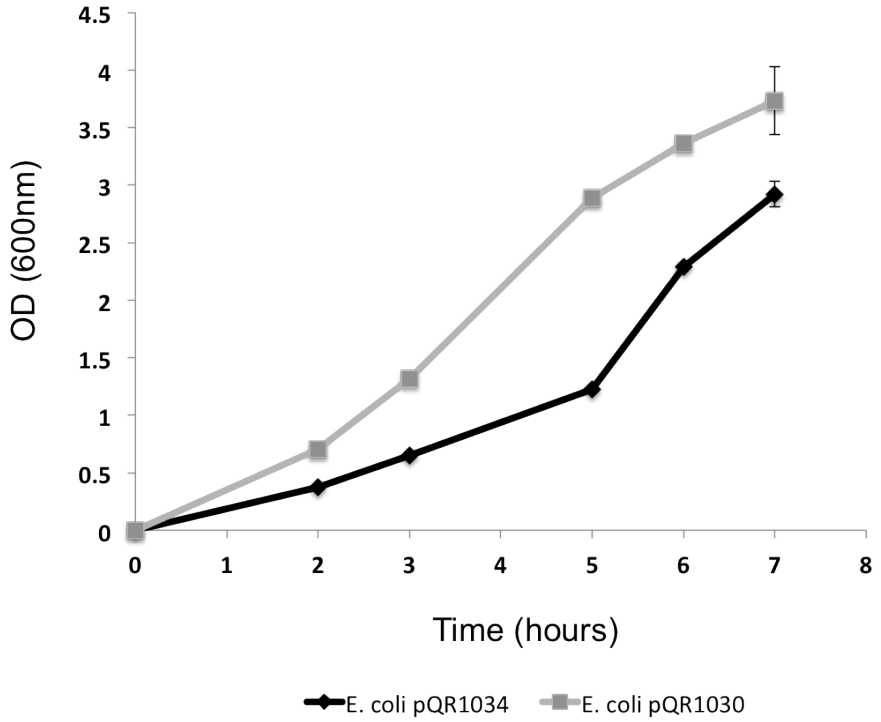


Figure 56- Growth profiles of the *E. coli* strains containing the expression vectors pQR1030 and pQR1034. The μ_{max} values were 0.787 h^{-1} and 1.068 h^{-1}, respectively. Error bars refer to the standard deviation of the experimental setup, calculated from triplicates of the same assay.

Another parameter that could be affected by the increased expression of CHMO was the stability of the biocatalyst inside the cell. In most of the NADPH oxidation assays, the lysate samples had been collected 8 hours after incubation, in a period when bacterial growth is transitioning from exponential to stationary phase. This is the period in which cellular density has reached its maximum. As the fermentation progresses into the stationary phase, several nutrients in the medium that are used as a carbon source become limiting and bacterial cells shift the metabolic machinery from protein expression to stress-induced survival, thus stopping bacterial growth and heterologous protein expression. Under these conditions, the active population of over-expressed heterologous proteins is expected to decrease, as a combined result from the decreased metabolic production and the cell-mediated degradation of proteins into constituents that can be used as raw material for other metabolic processes.

The results from Figure 57, obtained from NADPH oxidation assays conducted on pQR1034 *E. coli* lysates after 8 and 24 hours of incubation, showed

that after 24 hours of bacterial growth in shake flasks, in which no extra carbon source was added, the activity of CHMO was dramatically reduced in both strains. This clearly indicated that the biocatalyst was not stable inside bacterial hosts after a prolonged period of stationary growth. On the other hand, pQR1034 still exhibited a higher biocatalyst expression after 24 hours of growth, from which it could be concluded that the increased expression induced by the rplK promoter was not detrimental to the stability of CHMO.

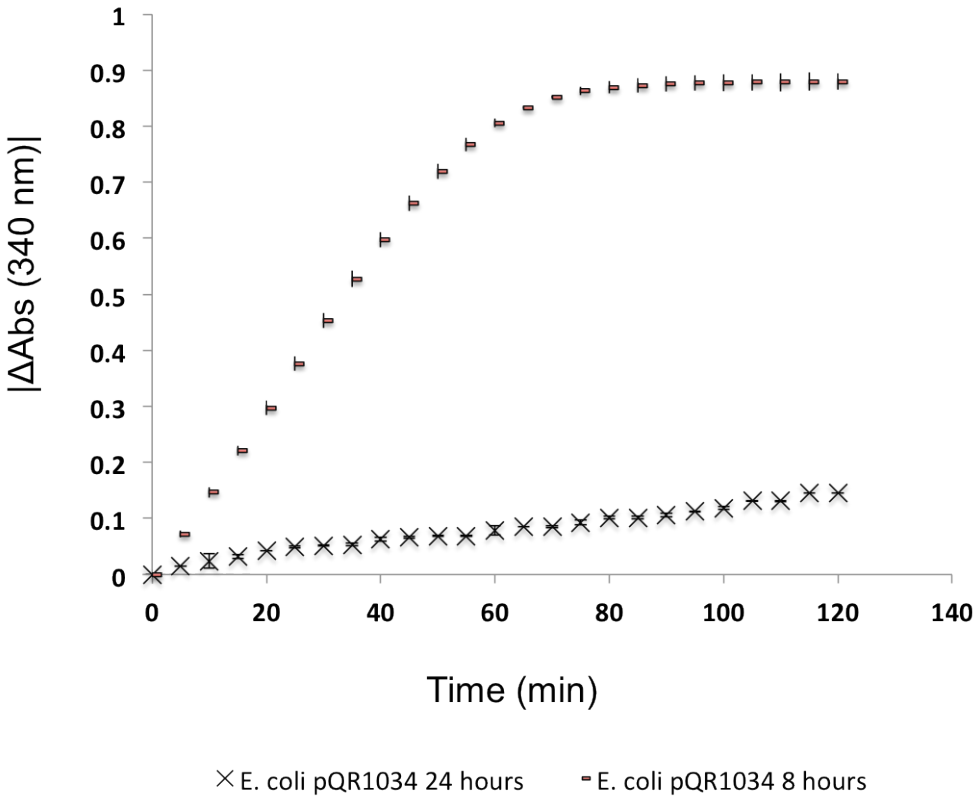


Figure 57- NADPH oxidation profiles of the pQR1034 *E. coli* lysates collected after 8 and 24 hours of bacterial growth. The specific enzymatic activity was reduced from 88 µmol min^{-1} $gDCW^{-1}$ to 2 µmol min^{-1} $gDCW^{-1}$ after 24 hours of incubation. Error bars refer to the standard deviation of the experimental setup, calculated from triplicates of the same assay.

To check if the increased protein expression resulted in the production of inclusion bodies, SDS-PAGE protein gels of the supernatant and pellet fractions from lysed bacterial cultures after 8 hours and 24 hours were performed (see chapter 2, section 2.3.3). Unfortunately, the results (Figure 58) did not show a significant difference in expression profiles between cultures containing the vector

pQR239, pQR1034, and an *E. coli* strain that did not express CHMO. The inability of the SDS-PAGE to differentiate between the expression levels of pQR1034 and pQR239 might be due to the fact that CHMO concentration inside the cells did not account for a large proportion of total protein content and therefore could not be discerned from other metabolic proteins of similar molecular weight. Regardless, the NADPH activity results clearly indicate that the increased protein expression from pQR1034 is not detrimental to the stability of CHMO inside *E. coli* TOP10, and therefore it was assumed that any increase in inclusion body formation would be counteracted by increasing levels of active biocatalyst.

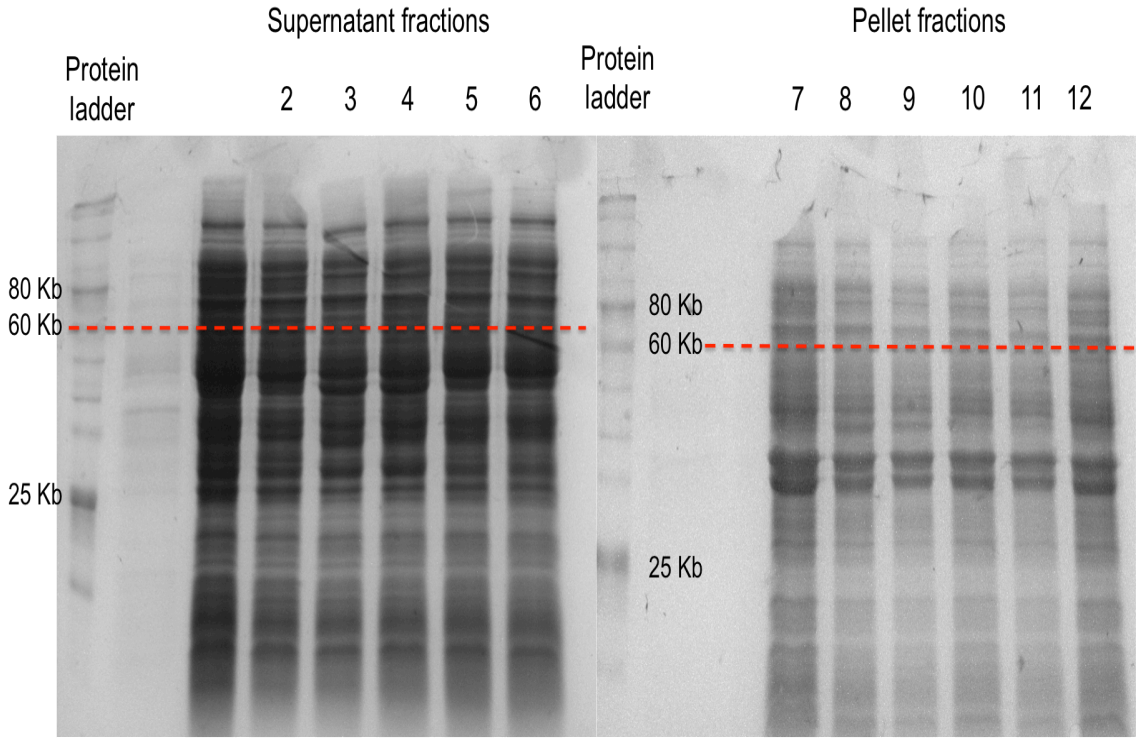


Figure 58- SDS-PAGE gel with fractions from *E. coli* cultures containing the expression plasmids pQR1030, pQR1034 and a control that did not express CHMO. Lanes 1 to 6 correspond to the supernatant fractions of lysates from the cultures, while lanes 7 to 12 correspond to the pellet fractions. Lanes 1, 2, 7 and 8: *E. coli* control; lanes 3, 4, 9 and 10: *E. coli* pQR1030; lanes 5, 6, 11 and 12: *E. coli* pQR1034. The red traced line indicates the 60 Kb threshold, below which one would expect to see over-expression of the 59 Kb biocatalyst.

5.2- NADPH Oxidation Profiles of *S. carnosus* TM300 Transformed with pQR1034.

The gentamicin resistance assays with *S. carnosus* TM300 containing the vectors pQR1032 and pQR1033 indicated that the introduction of the smaller rplK and rplJ promoters had resulted in 7 times increase in protein expression level compared to the initial construct pQR1030 (see chapter 4, section 4.9). As a result, similar results were expected when testing for CHMO activity in staphylococcal strains transformed with pQR1034. NADPH oxidation assays were performed on lysates of *S. carnosus* using the same reaction protocols as those used in previous biocatalytic assays. Despite the use of several different protocols for the production of lysates, no biocatalytic activity was detected on the supernatant fractions of the staphylococcal lysates (Figure 59). These results suggested that either CHMO was not being expressed or that the lysis methods used were detrimental to biocatalytic activity

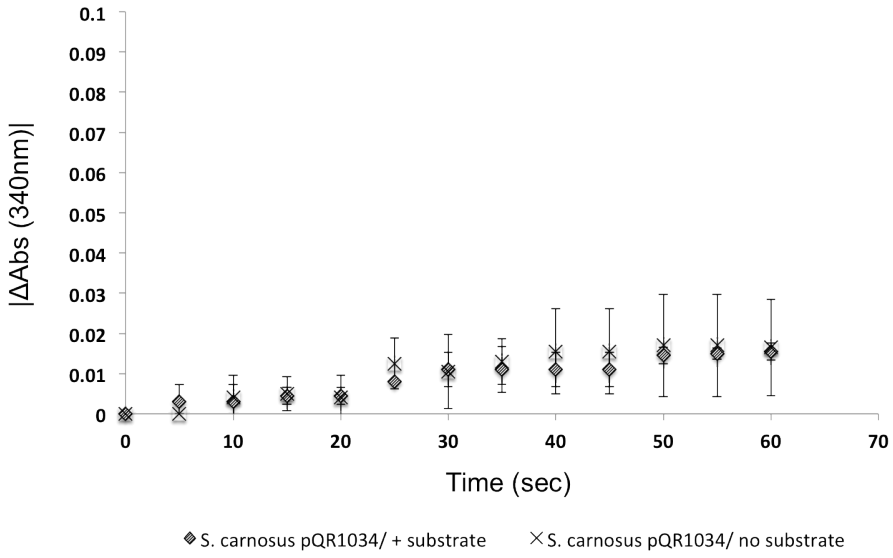


Figure 59- NADPH oxidation profiles of lysates from *S. carnosus* cultures containing the expression vector pQR1034 in the presence and absence of the substrate cyclohexanone. The y axis expresses the absorbance gradient between different time points of the reaction as an absolute number. Error bars refer to the standard deviation of the experimental setup, calculated from triplicates of the same assay.

In fact, the use of mechanical disruption methods did not produce efficient breakage, and required the use of extreme shearing conditions. Efficient sonication of cells was only achieved using a frequency of 20 KHz for periods of 30 seconds, which not only generated high levels of shear stress but also resulted in an increase of the overall sample temperature. Samples were sonicated in an ice water bath to prevent overheating, but even with this precautionary measure, the temperature of samples still rose above 37 ^{0}C during the sonication process. French Press induced cell lysis is a less stressful alternative to sonication that relies on cell disruption by pressure gradients, and thus generates less shear stress on cells. In addition, the cylinder cell that holds the sample is pre-chilled, preventing any over-heating that might occur during the disruption process. As a result, French press homogenization was the preferred mechanical lysis method to minimize any stress to the proteins in the supernatant that could kill CHMO activity. However, using the French Press to lyse the samples at a pressure of 1500 bar resulted in a very low degree of disruption when compared to the sonication protocol (data not shown). The supernatant obtained from this procedure also did not exhibit any biocatalyst activity, which suggests the lack of CHMO activity was due to translation or transcription problems, rather than detrimental effects from the lysis steps.

Alternately, staphylococcal samples were also treated with a lysostaphin containing buffer prior to mechanical lysis in order to weaken the cell wall structure, and consequently facilitate lysis. Similarly to results obtained with the other protocols, the supernatant fractions from samples disrupted in this manner did not exhibit any CHMO activity.

Together, the results from the NADPH oxidation assays hinted at the possibility that either CHMO was not being correctly expressed in *S. carnosus* TM300, or that it is not intracellularly stable. Alternately, another explanation for the lack of activity in the supernatant fraction of *S. carnosus* could be that the assay developed for measuring CHMO activity through NADPH oxidation raters was not applicable to the staphylococcal system. This possibility was unlikely, since reaction buffer used in the assays is optimal for activity of the CHMO, and the ratio in which the buffer was added to the supernatant fractions (4 volumes of buffer to 1 volume of supernatant) would be sufficient to dilute and negate any detrimental

effects that the supernatant environment might have on the activity of CHMO. On the other hand, there was very little information about proteins natively expressed in *S. carnosus* that could compete with CHMO for binding to NADPH, or that could work as a efficient NADPH recycling system.

In order to access if the results obtained in the oxidation assays were due to shortcomings in the methodology, or were the result of inherent limitations of the *S.carnosus* host, whole-cell biocatalytic studies were conducted on both *E.coli* and *S. carnosus* strains containing PQR1034. Analyzing the activity of CHMO in growing cells would give a clear indication of whether the biocatalyst was not being expressed at all, or if it had a very short half-life upon translation and folding.

5.3- Whole-Cell Biocatalysis of *E. coli* and *S. carnosus* containing pQR1034.

Most of the enzyme activity data produced for the expression vector pQR1034 had been obtained from measuring the oxidative activity of biocatalyst in lysates, which while being a fast and reliable assay to measure for activity, does not represent the conditions in which CHMO would be operating in a whole-cell biocatalytic process. As a result, in order to get a clear picture of how the biocatalyst expression profile of pQR1034 would affect the efficiency of whole-cell biocatalysis, whole-cell biocatalytic assays were performed on vector containing bacterial strains. In addition, as stated above, performing whole-cell biocatalysis with growing *S. carnosus* cells would give a stronger indication of whether the CHMO was expressed and stable in the intracellular environment.

Whole-cell biocatalysis was done using *E. coli* and *S. carnosus* strains in both active growth and in the resting state. For biocatalysis with growing cultures, colonies of the bacterial strains were inoculated in 5 ml of NBn2 and grown overnight, and subsequently transferred to 500ml shake flasks containing 100ml of working volume of NB2 medium with the respective selective antibiotic and 20 mM of the substrate cyclohexanone. The inoculated flasks were incubated for 24 hours at 37 ^{o}C and 250 rpm, with samples being taken at specific time points during the

fermentation. These samples were subsequently used for the GC analysis of cyclohexanone and ε-caprolactone concentrations present in the supernatant of grown cultures (see chapter 2, section 2.3.6). Results from the whole-cell biocatalysis of growing cultures were expressed as substrate and product conversion rates.

From the NADPH oxidation assays, it was expected for the *E. coli* TOP10 strain containing the vector pQR1034 to exhibit a biocatalytic activity profile that was linearly correlated to the strain's growth profile. The results from Figure 60B show that this prediction was correct. The highest rate of substrate conversion was observed during the exponential phase of bacterial growth, in which intracellular protein production was at its highest level. Similarly, as bacterial growth reached the stationary phase, the substrate conversion rate was also dramatically decreased, but not halted completely. This result re-iterated the hypothesis that CHMO has a short half-life inside of *E. coli* strains.

Alternatively, the decrease in the conversion rates could be due to product inhibition, as it has been previously reported that increasing concentrations of caprolactone inhibit CHMO activity (Doo et al. 2009). However, the total caprolactone concentration generated from the biocatalytic reaction did not reach the threshold at which it starts being inhibiting to cyclohexanone conversion.

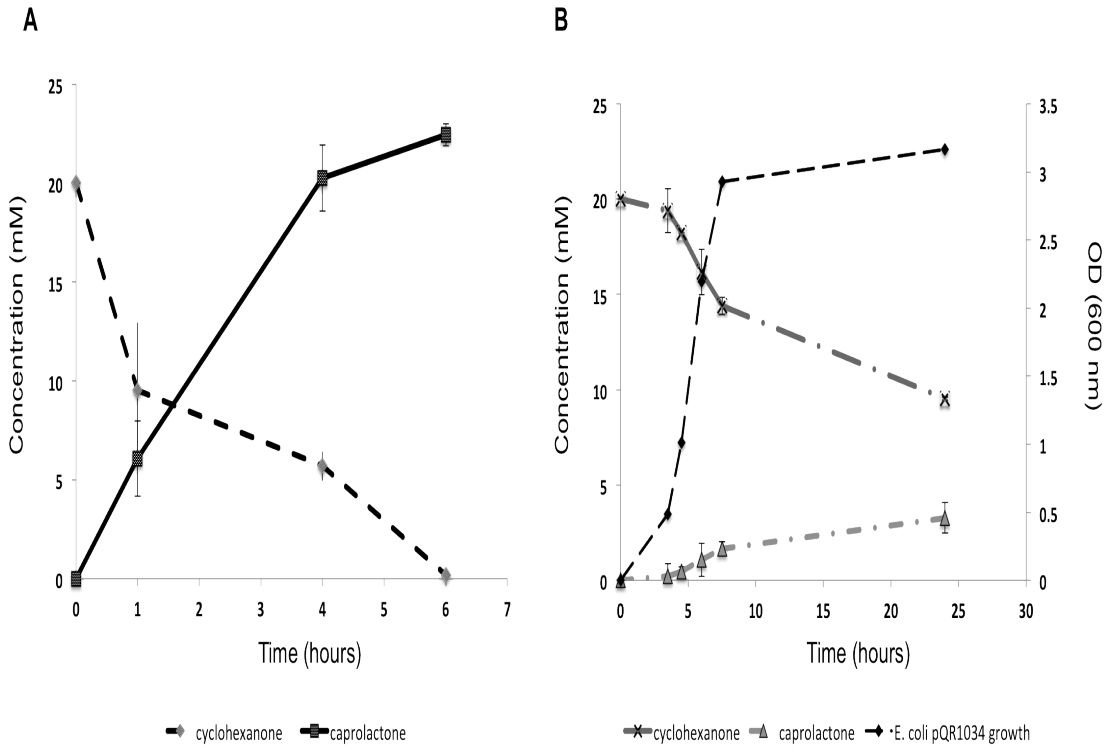


Figure 60- Cyclohexanone and caprolactone concentration levels in whole-cell biocatalytic reactions performed on *E. coli* cultures containing the expression vector pQR1034, in the resting state (A) and actively growing (B). The growth profile cultures performing bioconversion under active growing conditions were also included in the plot chart. Initial concentration of cyclohexanone was 20 mM. Error bars refer to the standard deviation of the experimental setup, calculated from triplicates of the same assay.

By contrast, the GC analysis of samples from the biocatalytic reaction using the *S. carnosus* strains showed that no substrate conversion took place during the 24 hours incubation (Figure 61B). The slight decrease in cyclohexanone concentration observed in the first hours of the reaction can be attributed to either measurement errors or evaporation of the substrate cyclohexanone, which has a low boiling point. The inability of the *S. carnosus* strain to convert the substrate corroborated the results from the NADPH oxidation assays, and hinted at several possible limitations of the S. carnosus pQR1034 biocatalytic system

One possible limitation would be for the biocatalyst to have a very short half-life inside the staphylococcal strain. If CHMO was actively expressed but suffered from a short half-life inside the cells, a small level of caprolactone conversion should still be detected in the first hours of the fermentation. However, the thick peptidoglycan cell wall might also work against cyclohexanone conversion by

imposing a barrier to the diffusion of the substrate into the intracellular space. A slow diffusion rate across the membrane coupled with the short half-life of the biocatalyst could account for the absence of any product formation.

Alternatively, the expression vector pQR1034 might not be stable inside the staphylococcal cells, even under antibiotic selection, thus preventing CHMO transcription. This possibility seemed unlikely, since both precursor vectors pQR1032 and pQR1333 still granted the staphylococcal host with a gentamicin resistant phenotype, in spite of any possible genetic instability. Unfortunately, the data obtained from the biocatalysis experiments was not sufficient to clear picture of what factors might be contributing to the lack of CHMO activity in *S. carnosus*.

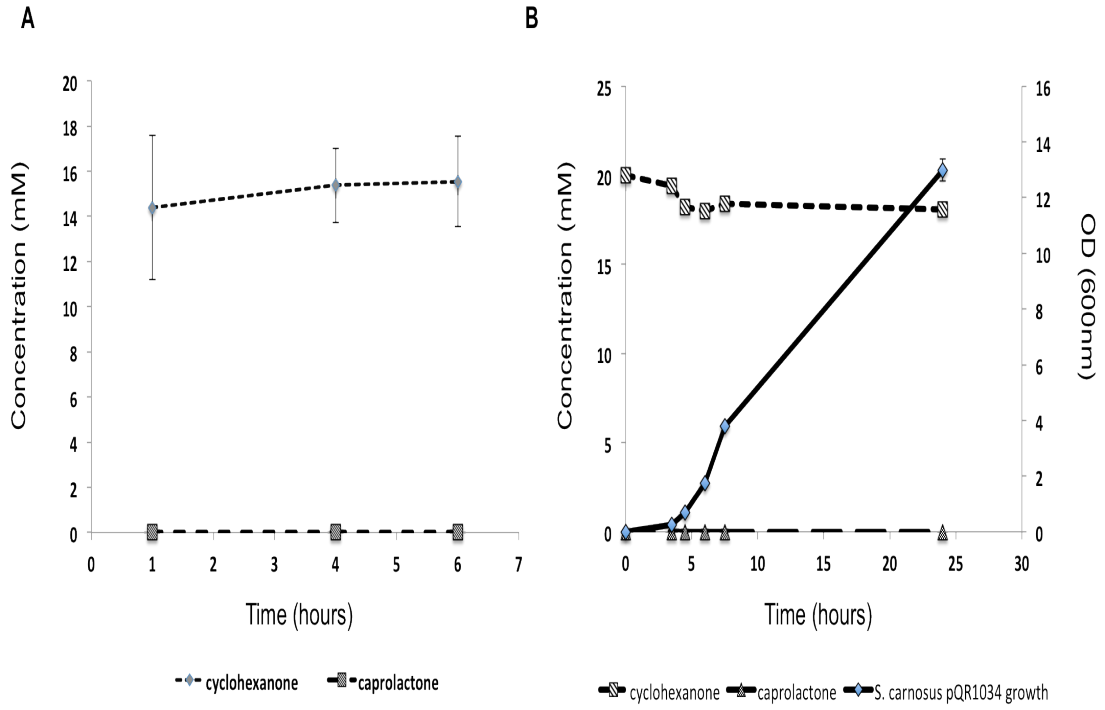


Figure 61- Cyclohexanone and caprolactone concentration levels in whole-cell biocatalytic reactions performed on *S. carnosus* cultures containing the expression vector pQR1034, in the resting state (A) and actively growing (B). The growth profile cultures performing bioconversion under active growing conditions was also included in the plot chart (B). Error bars refer to the standard deviation of the experimental setup, calculated from triplicates of the same assay.

The resting state biocatalysis was performed using bacterial cultures grown under similar conditions to those described above, with the exception that cells

were grown in the absence of the substrate cyclohexanone and collected after 8 hours of growth at the point of transition from the exponential to stationary phases. Cells were subsequently resuspended in a tenth of the initial volume of reaction buffer (50 mM phosphate buffer, pH 7) containing 10 g/L glycerol and 20 mM cyclohexanone. The resulting reaction samples were re-incubated for 8 hours at 30 ^{o}C and 250 rpm, with samples being taken at time points throughout the incubation for subsequent GC analysis. Figure 60A shows the results obtained from this biocatalytic process. Reactions done with the pQR1034-containing *E.coli* TOP10 strain achieved total conversion of cyclohexanone after six hours of incubation, with a maximum specific whole-cell CHMO activity of 8.7 µmol of caprolactone min^{-1} g DCW^{-1}. Since NADPH consumption is stoichiometrically linked to caprolactone production, with a 93% to 100% molar yield of caprolactone on NADPH consumption (O'Sullivan et al. 2001), the activity obtained from this assay could be compared to the NADPH oxidation rate obtained from the NADPH oxidation assays with the same bacterial strain, which was calculated to be 87.8 µmol min^{-1} g DCW^{-1} This comparison revealed an 10 times activity loss when biocatalysis was performed in the whole-cell system, which also had been reported for the pQR239 whole-cell biocatalytic system (Doig et al. 2003).

S. carnosus cultures used in the resting-cell biocatalysis did not produce any lactone product during the 8 hours of reaction (Figure 61A). This was to be expected from the previous assay using actively growing cells. SDS-PAGE gels of these samples did not indicate overexpression of the CHMO protein in either the supernatant or pellet fractions, further suggesting that the biocatalyst was not being expressed in the cells or only at very low level. However, these results were inconclusive and unreliable, as no overexpression could be detected in the SDS-PAGE gels of *E.coli* samples as well.

Faced with the inability to generate biocatalytic activity in *S. carnosus* TM300 containing pQR1034, it was decided that the best course of action would be to focus on some of the possible factors that could be responsible for this lack of activity.

5.4- Toxicity of Cyclohexanone and Caprolactone on *E.coli* TOP10 and *S. carnosus* TM300

One of the possible factors affecting cyclohexanone bioconversion in *S. carnosus* is the permeability of the cell wall to the substrate. The affinity of a chemical compound to the cell's lipid membrane and to the extracellular medium dictates its rate of diffusion across the cell wall into the intracellular space, where the biocatalytic process is performed. In many biocatalytic processes the cell wall becomes the main barrier to substrate accessibility, and therefore its affinity to the lipid membrane becomes a major limiting factor in the efficiency of whole-cell biocatalysis (Ni & Chen 2004). However, a compound's affinity to the lipid membrane is also directly related to its toxicity. Lipophilic compounds will inherently have a tendency to accumulate in the lipid membrane, which in extreme cases can cause pore formation, as well as loss of membrane function and structure (Sardessai & Bhosle 2002; de Bont 1998; Leon et al. 1998). Thus, the toxicity of a compound can be used indirectly as a measure of membrane permeability to that compound.

Several cyclic ketones used in CHMO-mediated oxygenation have been previously reported to be moderately toxic to *E. coli* biocatalytic process (Doo et al. 2009; Hilker et al. 2008) due to their lypophilic nature. To test if the permeability of staphylococcal cell wall to these compounds was comparable to *E. coli*, the tolerance of *E.coli* TOP10 and *S. carnosus* TM300 to cyclohexanone and caprolactone was measured by growing both strains in the presence of inhibitory concentrations of these compounds. Similar tolerance profiles between the strains would suggest that they have similar permeability to the compounds involved in CHMO biocatalysis.

For this purpose, colonies we cultured overnight in 5 ml of NBn2, and the resulting cultures were transferred to shake flasks containing a working volume of 100ml. The shake flasks were incubated for 8 hours at 37 ^{o}C and 250 rpm, and OD

measurements (600nm) were taken at specific time points during the incubation. In an first set of experiments, different concentrations of the cyclohexanone precursor, cyclohexane (C_6H_{12}), were added at the beginning and during shake flask incubation. Cyclohexane is a cyclic compound with a log P of 3.4, which lies within the window of toxic compounds that can diffuse and accumulate in the lipid membrane. In fact, cyclohexane is more aggressive to the cells than either cyclohexanone or the product caprolactone, and therefore using this chemical compound would give a good indication of difference in the permeability of the difference strains to this type of compounds. Figure 62 shows the results from the tolerance assays with cyclohexane. In the bacterial batches where cyclohexane was added at the beginning of the shake flask incubation, bacterial growth was completely halted upon addition of 100mM of the cyclic compound until after 6 hours of incubation. Once cultures started growing, *S. carnosus* showed a faster recovery rate (0.25 h^{-1}) than its *E. coli* counterpart (0.08 h^{-1}).

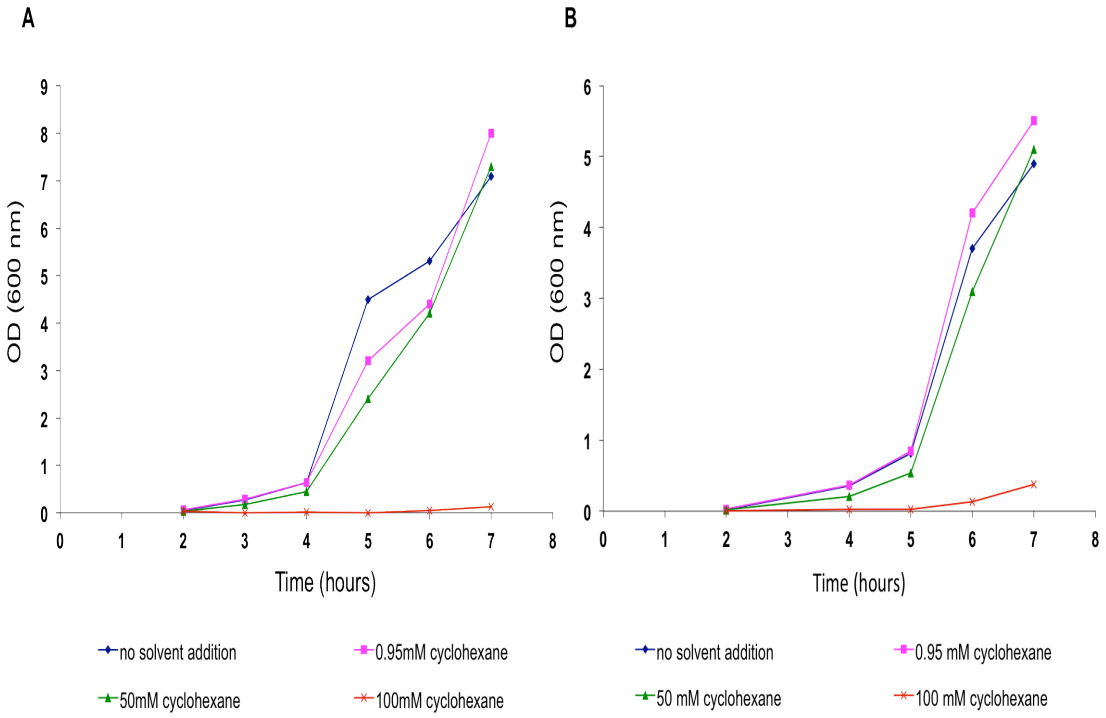


Figure 62- Growth profiles of *E. coli* TOP10 (A) and S. carnosus TM300 (B) cultures grown with increasing concentrations of cyclohexane, which was added at the beginning of incubation. Error bars refer to the standard deviation of the experimental setup, calculated from triplicates of the same assay.

This suggested that the Gram-positive bacterium is either less permeable to cyclohexane or it contains faster mechanisms of adaptability to the toxic effects of the cyclic compound. On the other hand, the different growth rates observed after hour 6 might have resulted from a decrease of cyclohexane concentration in the medium due to evaporation, together with the different inherent growth profiles of the two bacterial strains

The data obtained from the tolerance assays using lower concentrations of the toxic compound is more conclusive in showing the tolerance differences between the two strains, as in these cases bacterial growth is affected but not stalled. Indeed, as cyclohexane concentrations added were increased, a more pronounced decrease of the maximum growth rates of *E. coli* cultures was observed, from 3.86 h^{-1} to 1.96 h-1. By contrast, the maximum growth rate of *S. carnosus* cultures incubated with increasing concentrations of cyclohexane was not as severely affected. A concentration of 50 mM of the compound resulted in a small decrease of the maximum growth rate from 2.89 h^{-1} to 2.52 h^{-1}. Together, these results corroborated the suggestion that *S. carnosus* is more tolerant to cyclohexane, either through cellular mechanisms that allow for adaptability to toxic lipophilic compounds, or through differences in cell wall permeability to these compounds.

Similarly, adding an inhibitory concentration of cyclohexane to actively growing cells produced a stronger effect on the *E. coli* cultures. Cyclohexane was added after 3 hours of incubation, at the beginning of the exponential phase. At this time point, cells were in a state of high metabolic activity and consequently their ability to rapidly adapt to toxic compounds was greater than during the lag phase. Conversely, the addition of cyclohexane during the exponential phase had a stronger effect on bacterial growth of both strains than addition of the compound at the beginning of incubation (Figure 63).

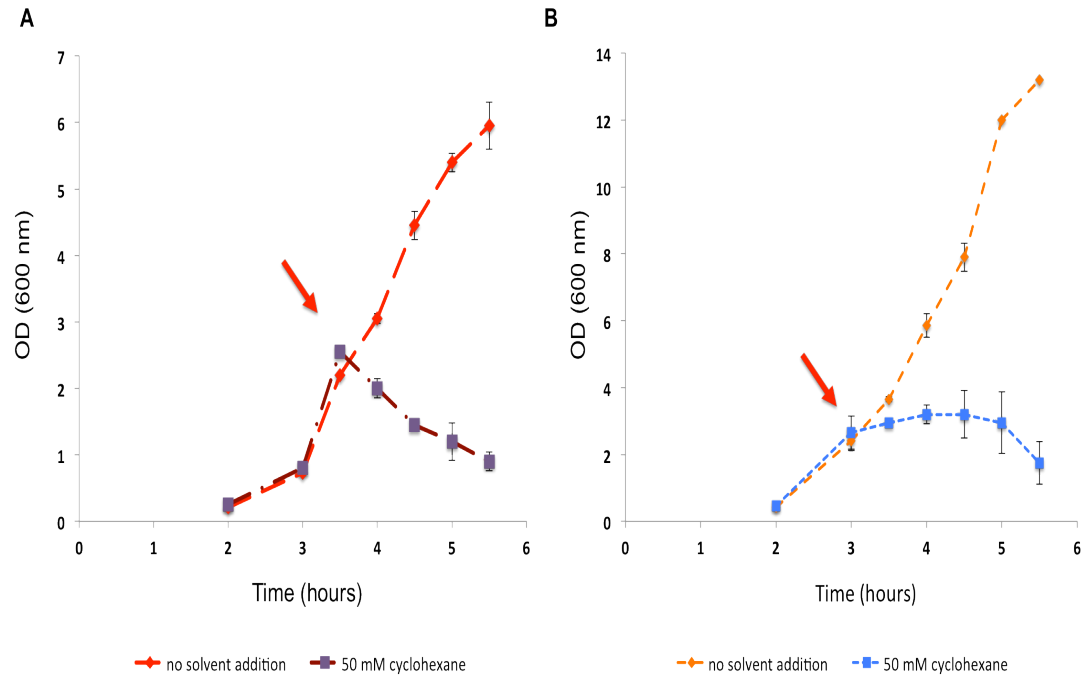


Figure 63- Growth profiles of E. coli TOP10 (A) and S. carnosus TM300 (B) cultures grown in the absence and presence of 50 mM of cyclohexane. The compound was added after cultures reached a specific OD (as indicated by the red arrow). Error bars refer to the standard deviation of the experimental setup, calculated from triplicates of the same assay.

Adding 50 mM of cyclohexane after 3 hours of incubation had a strong negative effect on *E. coli* growth, as its growth rate became negative after the addition. Conversely, cyclohexane had a stationary-phase-like effect on the growth of *S. carnosus* cultures by stalling the growth rate. However, a pronounced decrease in cell density was only observed after 5 hours of incubation, at which point cultures were transitioning from exponential growth to the stationary phase.
Alternately, a correlation could be drawn between the effect of cyclohexane and the overall growth of the bacterial strains. The fact that *S. carnosus* cultures showed higher growth rates than the *E. coli* counterpart under the experimental conditions of these assays could explain the lower rates cyclohexane-induced growth inhibition. Therefore, the increased tolerance of *S. carnosus* to cyclohexane could be the result of the better overall growth of this strain, instead of differences in permeability or adaptability mechanisms. This directly contradicted the results from experiments in which cyclohexane was added at the beginning of incubation, wherein the two bacterial strains exhibited similar growth rates.

It is worth noting that cyclohexane was not very soluble in aqueous solution, and consequently to facilitate homogeneous mixing of the latter 1% (v/v) of the surfactant X-Triton 100 was added to the cultures. At this concentration, the surfactant was shown to be toxic to cells on its own, and as a result the toxicity observed in the cyclohexane assays is in fact the combined effects of cyclohexane and X-Triton 100.

A second set of tolerance assays was done on the compounds directly used in the CHMO biocatalysis assays throughout this project, cyclohexanone and ε-caprolactone. With a log P value of 0.81 and 0.61, respectively, these compounds are water miscible, and therefore no surfactant was added to the cultures. In addition, these compounds are much less lipophilic than the precursor cyclohexane, and consequently less toxic. The assays were performed under similar conditions to those described for the cyclohexane tolerance assays, but the compounds were only added during the exponential phase of bacterial growth. Figure 64 shows the results obtained from these assays. Unlike the cyclohexane assays, strong inhibition was only observed around 2 hours after addition of the chemicals. Even so, the results showed that the *E. coli* strain was generally less tolerant to increasing concentration of the compounds than the staphylococcal counterpart. In particular, bacterial growth of *E. coli* was completely stopped after the addition of 100 mM of cyclohexanone, which was not the case for *S. carnosus*. In addition the disparity between the tolerance of the bacterial strains to cyclohexanone and ε-caprolactone could not be explained by differences in growth rates, which were comparable throughout the assays.

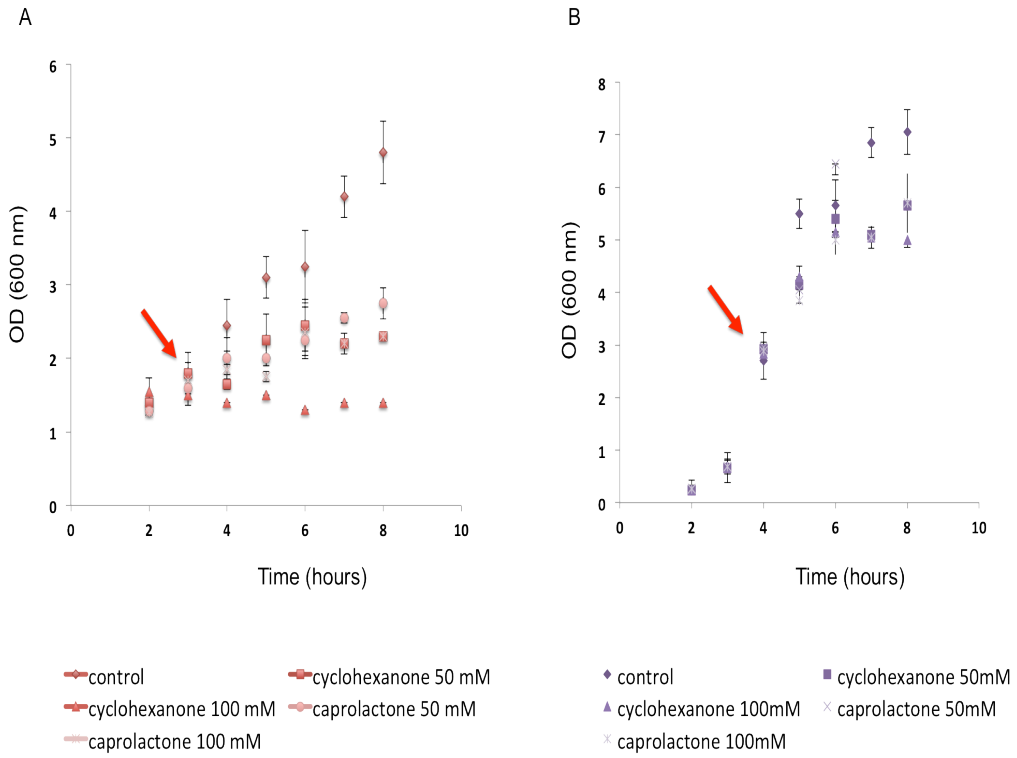


Figure 64- Growth profiles of E. coli TOP10 (A) and S. carnosus TM300 (B) cultures grown with increasing concentrations of either cyclohexanone or caprolactone. The compounds were added to cultures after reaching a specific OD (as indicated by the red arrow). Error bars refer to the standard deviation of the experimental setup, calculated from triplicates of the same assay.

The tolerance to caprolactone was higher in both strains than tolerance to cyclohexanone. This corroborates predictions based on the log P values of the compound, under which caprolactone is the less lipophilic compound, and therefore should not accumulate in the cell membrane as readily as cyclohexanone. Together, these results indicated that *S. carnosus* is more tolerant to the substrate and product from CHMO biocatalysis that had been used in whole-cell biocatalytic assays. This higher tolerance when compared to *E. coli* might be a result of either better mechanisms for neutralization or adaptation to the toxic compounds, or lower permeability to the compounds. However, the results also suggest that the cell wall of *S. carnosus* is not completely impermeable to the diffusion of cyclohexanone and caprolactone, otherwise no toxic effect would be observed on bacterial growth of this organism. However, these results don't rule out the possibility of low diffusion rates of these compounds across the

staphylococcal membrane being a limitation for CHMO-mediated whole-cell biocatalysis.

5.5- Plasmid Stability Assays with pQR1034

Another factor that could affect expression of CHMO in *S. carnosus* is the instability of the expression vector pQR1034 inside the staphylococcal strain. Transformation of hybrid plasmids into *S. carnosus* had been previously reported to be difficult due to the fact that the organism has mechanisms in place to degrade DNA it recognizes as foreign (Löfblom et al. 2007). Therefore, there is the possibility pQR1034 is genetically and structurally unstable in *S. carnosus* due to the Gram-negative portion originated from the pTTQ18 vector, which encompasses a large percentage of the vectors' genetic content.

Correspondingly, the total GC content of pQR1034 is 46.1%, which is around 12% higher than the native genome of *S. carnosus*. A similar argument could be made for the GenR expression vectors pQR1032 and pQR1033. However, the fact that these vectors allowed for the phenotypic expression of gentamicin resistance in *S. carnosus*, together with the fact that the chloramphenicol antibiotic used in all the biocatalytic assays as a selection factor were strong counter-arguments to vector instability being the reason for lack of CHMO activity. Nonetheless, this possibility was tested by growing *S. carnosus* cells containing the vector pQR1034 for 3 days under non-selective conditions.

Samples were taken at different time points throughout the process and plated on plates with and without chloramphenicol, which were subsequently incubated overnight. The percentage of plasmid loss during the incubation period was derived from the difference between the number of colonies grown on plates with antibiotic selectivity and the total number of colonies, which were grown on plates without antibiotic. These assays were conducted with *E. coli* pQR1034 and pQR1030, as well as the staphylococcal vector pCT20, which was used for comparison of a genetic system containing the same origin of replication as pQR1034. Cultures from the different strains were grown in 50 ml Falcon tubes

containing 5ml of NB2 medium, which was replaced every 24 hours by inoculating 1 ml of the overnight cultures into fresh media. No antibiotic selection was used throughout the incubation period.

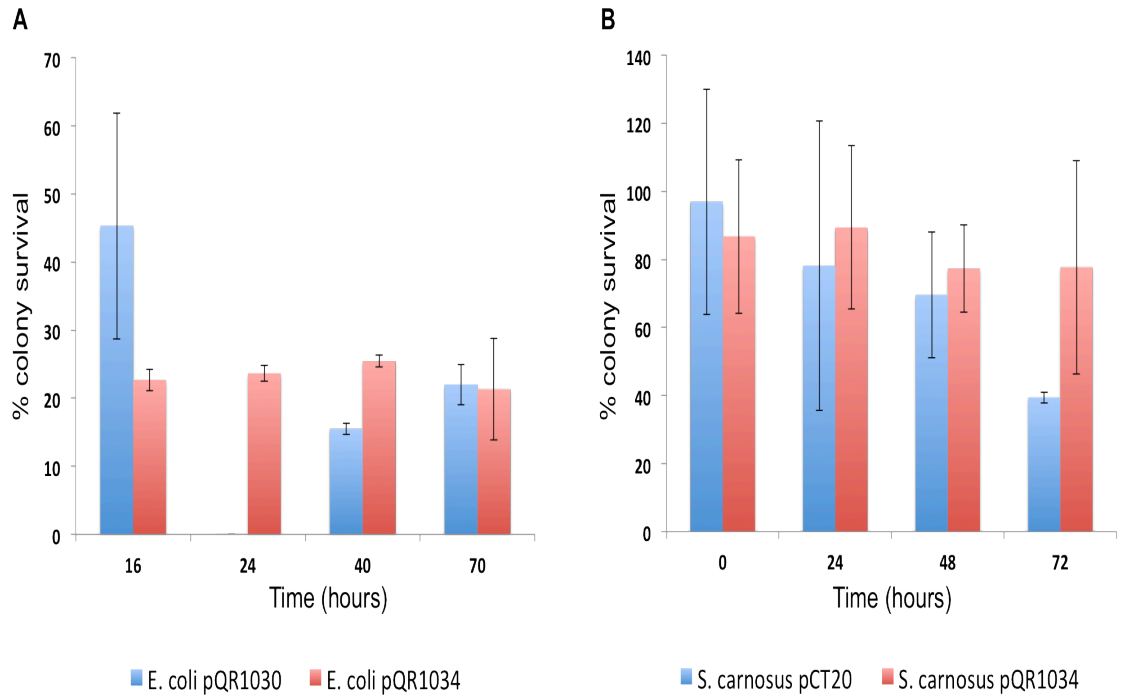


Figure 65- Plasmid stability assays with bacterial cultures incubated for 3 days in the absence of antibiotic selectivity. A) *E. coli* TOP10 cultures containing the expression vectors pQR1030 and pQR1034. B) *S. carnosus* TM300 cultures containing the expression vector pQR1034, as well as the staphylococcal plasmid pCT20. Error bars refer to the standard deviation of the experimental setup, calculated from triplicates of the same assay.

Figure 65 shows the results from the plasmid stability assays with the different bacterial strains. From these results, it could be observed that the vector pQR1034 was relatively stable in *E. coli*, with levels of surviving colonies fluctuating between 20 and 25 % of total number of colonies after 16 hours. If it was assumed that at the start of the experiment the total number of colonies would be equal to the number of colonies containing the vector, it could be subsequently concluded that there was a large plasmid loss in the first 16 hours of incubation. However, this initial instability cannot be attributed to overexpression of the biocatalyst CHMO, as the tightly controlled expression vector pQR1030 also exhibited a similar degree of plasmid loss during the first 16 hours of incubation.

Despite the initial drop, the levels of plasmid-containing colonies was maintained at around 20% of total number of colonies throughout the rest of the incubation period, suggesting that the plasmid constructs were stable in the *E. coli* strains after the first 16 hours of incubation.

Similarly, the results showed that the pQR1034 vector was stable in *S. carnosus* throughout the incubation period, with around 80% of the total number of colonies still containing the plasmid after 3 days of incubation. This stability outperformed results from assays conducted on the native staphylococcal vector pCT20, which was present in less than 40% of the total colony population by the end of the incubation. Therefore, from these results it could be concluded that the expression vector is stable in *S. carnosus* despite the presence of foreign DNA and the different GC content. As a result, the lack of CHMO activity in *S. carnosus* could not be explained by vector instability.

5.6- Trouble-shooting CHMO Expression in *S. carnosus*

Both the whole cell biocatalytic assays and NADPH oxidation assays suggested that CHMO was either not stable in the intracellular environment of *S. carnosus*, or that the gene was not being correctly transcribed or translated. Conclusions could not be drawn either way due to the fact that the supernatant and pellet fractions of samples loaded unto SDS-PAGE gels were inconclusive as to whether CHMO was present. This was even the case in the *E. coli* strains where a high degree of activity could be observed.

Correspondingly, one of the first and crucial experiments to be done in trouble-shooting the expression of CHMO in *S. carnosus* would be to perform western blots of the *E. coli* and *S. carnosus* strains containing the expression vector pQR1034. Western-blots are quantitative experiments that allows for the quantification of a protein fraction in a protein gel by using fluorescent antibody markers that bind to specific recognition sequences on the target protein. A common practice is to use an anti-poly histidine monoclonal antibodies that target

his-tags on proteins, and are attached to a horseradish peroxidase that allows for the subsequent detection of the bound protein by converting a soluble substrate into a visible product that accumulates around the area where the protein diffused in the gel. A linear correlation can then be established between the intensity of the visible signal and the amount monoclonal antibody body bound to the tag of the protein, which in turn is proportional to the amount of protein in the samples. Thus, Western blotting is a very important experiment to conduct, because it would allow not only for the detection of CHMO in the different fractions of *S. carnosus* cell lysates, but would also give an estimate of the biocatalyst concentration in these fractions.

To this effect, the CHMO gene in pQR1034 was replaced with an his-tagged homolog, which would serve as the marker for binding of anti-poly histidine monoclonal antibodies to the protein during the western-blot protocol. The positioning o f the his-tag in the C-terminus of CHMO has been shown previously not to be detrimental to its activity, and therefore introducing this tag should not interfere or change our previous characterization of CHMO activity in the different expression constructs. Unfortunately, the western-blot experiments could not be conducted due to time constraints, and therefore would have to be performed in future studies with the *S. carnosus* constructs.

Western blots of *S. carnosus* cultures containing pQR1034 would answer several questions. Primarily, they would allow for a direct assessment of the levels of pQR1034-mediated CHMO expression inside the staphylococcal strain. Subsequently, the results would enable the analysis of the stability of the biocatalyst in the intracellular environment by measuring the distribution of CHMO within the different fractions of lysate samples. Correctly folded and stable protein would be primarily present in the supernatant fractions, whereas aggregates would be partitioned into the pellet fractions. Inclusion bodies, which are highly dense miss-folded protein aggregates, would be mainly present in the insoluble fraction of lysates.

Western blots would be conducted not only on samples after 8 hours of incubation, but also on samples taken at time points throughout culture fermentation in order to test the theory that the CHMO is being actively degraded by native proteases. If this theory is correct, we would expect to see a low

peroxidase signal in the first hours of growth, followed by either a decrease in signal throughout the rest of the incubation period or a stable, non-fluctuating signal that corresponded to an equality between the rates of protein expression and the rates of proteolytic digestion. Depending on the results of the Western blots, several other experiments would be performed (Figure 66).

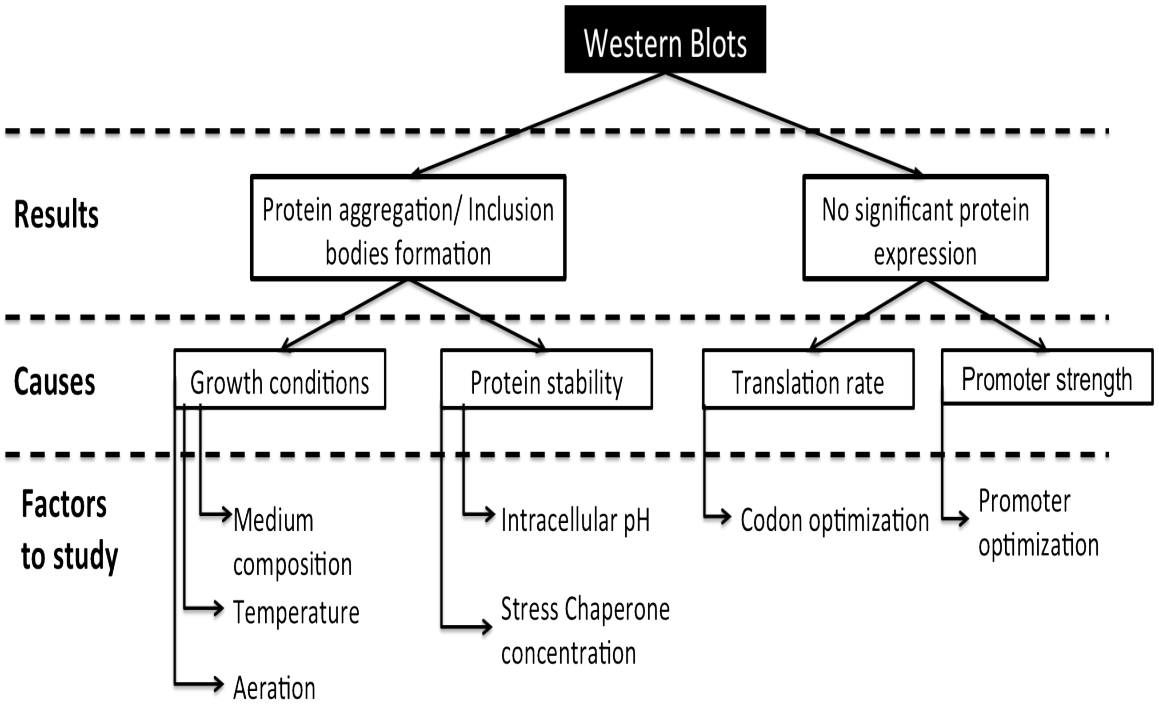


Figure 66- Diagram representation of the outcome of the western blot experiments and the factors that could be studied dependent on these out-comes.

One possible outcome of the western blots could be for most of the overexpressed CHMO to be either miss-folded or aggregated into inclusion body. This could be symptomatic of a high level of rplK-mediated constitutive expression, as the native chaperone activity would not be able to cover for the correct folding of increasing levels of heterologous proteins. A way to test the effect of native chaperone concentrations on the folding of the overexpressed CHMO would be to grow cultures at different temperatures, or after heat-shock treatment. Chaperones are housekeeping genes that are highly expressed in situations of stress to the native organism as a measure to protect the integrity and stability of native proteins. This increase in expression can be triggered by heat-shock events, i.e.

when there is a sudden temperature fluctuation, or by growing cultures in sub- or above- optimal growth temperatures. Since the temperature used in all growth and biocatalytic assays in this project was maintained constant at 37 ^{o}C, the effects of native stress response in *S carnosus* to the expression of CHMO has not yet been tested.

Correspondingly, one experiment to be conducted would involve using a low temperature in the initial lag stages of *S. carnosus* shake flask incubation, followed by a drastic shift of temperature during the exponential phase, which expectedly would trigger chaperone overexpression. Incubation of bacterial cultures at lower temperatures than 37 ^{o}C could also be conducted as a means to decrease the rates of protein over-expression, thus increasing the time window for the proteins to be properly folded. Temperatures such as 30 ^{o}C and 25 ^{o}C have been routinely used for *E. coli* cultures wherein heterologous proteins are highly expressed, or are unstable and consequently require a longer post-translation processing time to be folded properly.

The composition of media used for bacterial growth could also be a factor contributing to the levels of protein expression. Thus, another experiment could revolve around the effects of different media, such as NB2 or B2, and other more minimal media, on the expression of CHMO. Subtle differences in the medium composition, such as the salt concentrations, carbon source levels, and pH, change the manner in which cells react to the extracellular environment, and consequently their ability to express stable heterologous proteins.

For instance, the pH of external media might have a direct impact on the stability of intracellular biocatalysts by affecting the internal pH of cells. Studies done on lactic acid bacteria have shown that the intracellular pH (pHi) of these strains changes depending on the pH of the extracellular media, and in most cases the shift in pHi reflects similar shifts in the media.

More importantly, the faster growth rates and higher densities of *S. carnosus*, when compared with the *E. coli* counterpart, result in a faster depletion of oxygen levels during shake flask incubation, which was the preferred method for biomass production in this project. In turn, oxygen starvation promotes the production and excretion of acetic acid through anaerobic fermentation, which would consequently result in the acidification of the external medium. If *S.*

carnosus behaved in a similar fashion to lactic acid bacteria, the growth-induced acidification of external pH would correspond to a decrease in the pHi of growing cultures. The fact that the staphylococcal strain is tolerant to lower pH may partly explain why it can maintain high growth rates even in high levels of anaerobic fermentation. However, the decrease in the pHi would be detrimental to the activity of CHMO, which has been shown to greatly decrease at pH lower than 9 (Doig et al. 2003).

The incubation of staphylococcal cultures with buffered media at different pH would allow us to test this hypothesis. Conversely, there are methods for the direct measurement of pHi. One such non-invasive method relies on a GFP variant, ratiometric GFP (Olsen et al. 2002; Webb et al. 2001), which alters its excitation spectrum depending on the pHi. By replacing the CHMO gene in pQR1034 for this reporter gene, we would be able to test if, under the conditions used for the biocatalytic assays, the pHi of *S. carnosus* is significantly more acidic than the *E. coli* counterpart.

The acidification of the external environment during bacterial incubation was not helped by the limitations of the shake flask containers used for the incubations. The shake flask incubation experiments conducted throughout the project are involved the use of conical flasks that were completely sealed and consequently did not permit gas exchange between the exterior and the shake flask air pocket. Therefore, cultures growing using this apparatus had a limited supply of oxygen. The study of aeration is rather limited in conical shake flasks. Different flask geometries could be used to generate different oxygen mass-transfer rates between the air pocket and the growth media. Alternately, a more detailed analysis of the effects of aeration, including mass-transfer rates and oxygen concentrations, would require a fermentation reactor set-up, containing an impeller system for modulation of mixing and a gas input for modulation of oxygen concentration in the fermentation medium.

Figure 67 shows a DoE two-level factorial experimental design to test the effect carbon source concentration, growth temperature, heat-shock treatment, aeration, and external pH on the expression of CHMO in *S. carnosus*. This experiment would cover many of the factors that could have an influence of protein production, but they require the setting-up of 32 shake flask incubations. A

simplified and more efficient alternative would be to use 96-microwell plates, which have a proven record as tools for scale-down modeling (Baboo et al. 2012; Lye et al. 2003). In addition, a recent methodology has been developed to allow for the non-invasive measurement of oxygen levels in microwells (Ferreira-Torres et al. 2005), thus increasing the value of microwell systems as a better alternative to shake flask experiments As a result, the whole experiment set-up could be performed using 96 well plates with different wall conformations and incubated at different temperature, thus saving the time-consuming an cumbersome procedure to prepare and manipulate 32 conical shake flasks.

All the experiments discussed above have been designed with the assumption that CHMO was being overexpressed but not stable in the intracellular environment of *S. carnosus*. On the other extreme of the hypothetical outcomes, the Western blot results could indicate that very little or no biocatalyst expression was occurring during incubation. In this case, both the strength of the expression vector and the codon bias of the protein become the most important factors to trouble-shoot.

The gentamicin resistance experiments have suggested that both small rplK and rplJ promoters allow for a high level of GenR constitutive expression (see chapter 4, section 4.11), and thus expression pQR1034 should have a similar expression profile. However, it is possible that the resistance to gentamicin is not linearly correlated with the levels of GenR, and that low levels of the enzyme are enough to produce a strong response against the antibiotic. One way to test this theory would be to optimize gene expression, either by changing the staphylococcal promoter in pQR1034 for those already described in the literature, or by conducting another round of genome mining for strong constitutive promoters. In the next section of this chapter, we discuss in more detail this optimization.

Codon bias also becomes a potential trouble-shooting factor due to its effect on protein translation. The codon bias of a protein gene sequence becomes rate limiting in translation when codons that are rare in the host organism are employed repeatedly in the sequence, since the concentrations of the available tRNA for these codons are low in comparison with codons that are used more routinely in the host cells. Consequently, long strings of rare codons in gene sequences would

greatly inhibit the rates of translation. Since the CHMO is not native from *S. carnosus*, a valid line of enquiry would be to test if the optimization of the codon bias of CHMO to fit the codon preferences of the staphylococcal host would improve the expression of active proteins.

		Level	
Factor	Symbol	-	+
Glucose concentration (g/L)	A	0.5	2
Growth temperature	B	25	37
Heat-shock	C	Yes	No
Oxygen Levels (baffles)	D	No baffles	Baffles
Medium pH	E	5.5	7.5

	A	B	C	D	E
1	-	-	-	-	-
2	+	-	-	-	-
3	-	+	-	-	-
4	+	+	-	-	-
5	-	-	+	-	-
6	+	-	+	-	-
7	-	+	+	-	-
8	+	+	+	-	-
9	-	-	-	+	-
10	+	-	-	+	-
11	-	+	-	+	-
12	+	+	-	+	-
13	-	-	+	+	-
14	+	-	+	+	-
15	-	+	+	+	-
16	+	+	+	+	-
17	-	-	-	-	+
18	+	-	-	-	+
19	-	+	-	-	+
20	+	+	-	-	+
21	-	-	+	-	+
22	+	-	+	-	+
23	-	+	+	-	+
24	+	+	+	-	+
25	-	-	-	+	+
26	+	-	-	+	+
27	-	+	-	+	+
28	+	+	-	+	+
29	-	-	+	+	+
30	+	-	+	+	+
31	-	+	+	+	+
32	+	+	+	+	+

Figure 67- DOE design for the study of the effects of five factors on the expression of CHMO simultaneously. The factors A to E correspond to glucose concentration, temperature of incubation, heat-shock of cultures before incubation, the use of different geometries of 96-well plates (oxygen levels), and the pH of the medium, respectively. Each factor would be explored at two levels, high (+) and low(-). The results of the DOE experiment would be expressed in amount of protein expressed per gram of DCW.

In conclusion, the execution of the experiments described above would depend upon the results from the initial Western blot. Without the input that this experiment would give, we would have multiple directions in which to tackle the CHMO expression problem but no clear direction as to which would be the correct approach. It is worth noting that all the experiments described in this section were designed using my current empirical knowledge about the *S. carnosus* expression system, and therefore are not by all means an exhaustive list of all possible experimental studies that could be conducted. Rather, they are the experiments we found to be the most pertinent in the context of answering the question of CHMO expression in *S. carnosus*.

5.7- Optimization of the pQR Expression System

One factor that could be directly related to the lack of CHMO expression in *S. carnosus* is the strength of the staphylococcal promoters in the pQR expression systems. These promoters were either taken directly or iterated from putative promoter regions for highly expressed genes in the genome of *S. carnosus*, and therefore assumed to be natively useful constitute protein expression. Accordingly, the gentamicin resistance assays suggested that this was indeed the case for the smaller re-designed promoter variants (see Chapter 4, section 4.11). However, we did not have any means to compare the performance of these promoters with previously reported systems such as the lipase promoter or the xylose promoter, and therefore categorization of promoter strength for the promoters used in this project was subjective at best. Moreover, the bank of promoters used in the project is too small to reach any conclusions about their optimization in *S. carnosus*. We only tested three sequences, two of which were variants of the same promoter. Following the underlying principle behind the construction of the vector pQR1030, it should have been possible to study a much larger number of genomic promoters, which would provide a better understanding of how different sequences affect

promoter strength, but also a larger database of expression systems from which to compare the strengths of individual promoters.

Therefore, one approach to optimizing expression of CHMO in pQR1034 would be based on the random selection of a large number of putative promoter sequences from the genome of *S. carnosus*. Using the ability of the pQR1031 reporter vector for promoter interchangeability, large number of genomic sequences could be screened for promoter activity, thus creating a database of both constitutive and regulated native promoters from which an optimized expression variant could be more easily identified. One advantage of this approach is that it bypasses the need to use statistical *in-silico* models for the identification of genomic pockets of potentially highly expressed genes that have not been confirmed for *S. carnosus*. However, this shotgun approach to promoter screening would require extensive trial-and-error studies that would be both time and resource consuming.

Another approach to promoter optimization would be analogous to methodologies used in this project, in which sequences upstream from HEGs are re-designed into smaller promoter units. Several of these were designed in this project but not tested (Figure 72). As highlighted previously, the creation of smaller, more concise versions of the promoter regions upstream of putative highly expressed genes has the advantage of filtering out sequences of unknown regulatory functions, thus creating promoter units that are capable of inducing constitutive transcription. However, the constitutive nature of these units can become detrimental to the overall expression and stability of heterologous proteins, either by inducing the formation of inclusion bodies or by triggering stress-related proteolysis.

Therefore, a more efficient approach would be to design hybrid synthetic promoters that combined native genomic promoter elements from *S. carnosus* with regulatory sequences that have been previously characterized (Figure 68). Recently, software for the comparison and inference of genomic regulons called RegPredict (http://regpredict.lbl.gov/regpredict/) has assembled together databases of DNA binding sites for known transcription factor involved in the regulation of metabolic pathways, including several palindromic sites identified in staphylococcal species that could be used to moderate the expression of strong

genomic promoters. By combining the sites recognized in this database, which are allocated specifically to particular regulatory pathways, with the smaller sequences identified from the upstream promoter regions of HEGs, a comprehensive catalogue of inducible staphylococcal promoters could be constructed. Small promoters with high constitutive activity could be combined with different regulatory elements in order to modulate the method of regulation to fit different biocatalytic requirements.

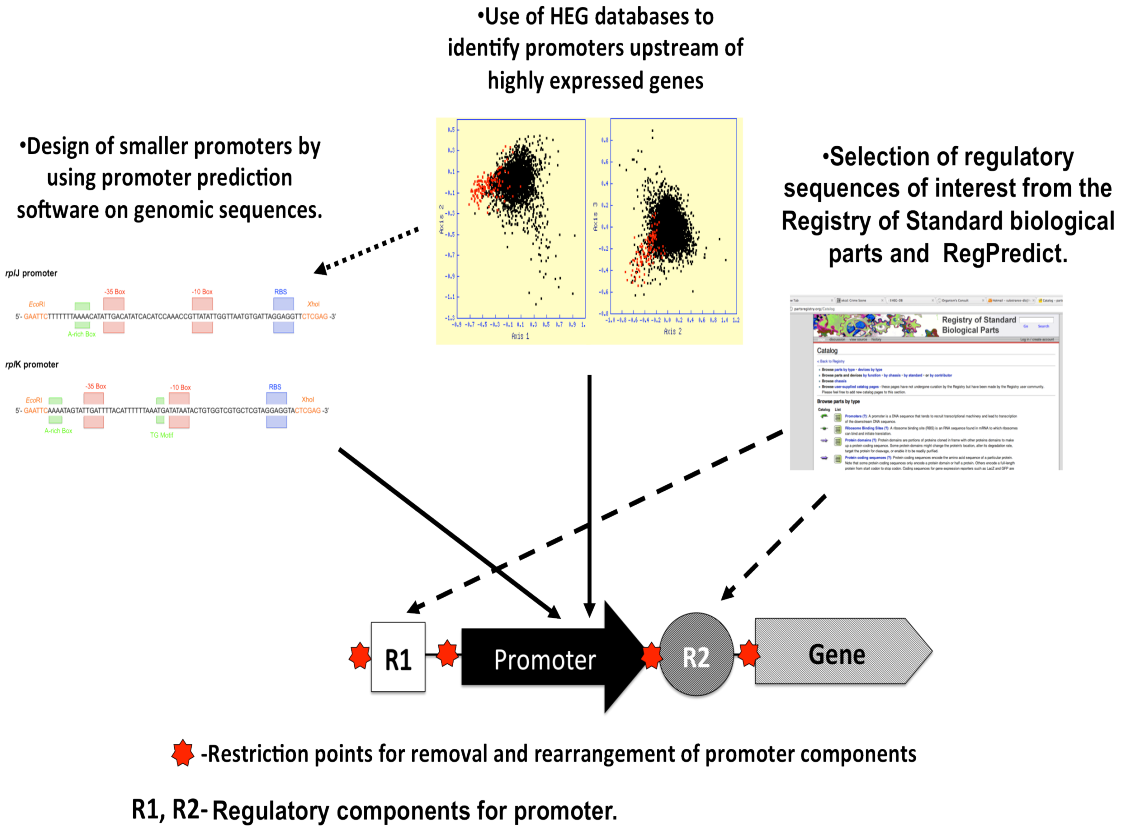


Figure 68- Representation of the different strategies used for the optimization of the promoter system for heterologous protein expression in *S. carnosus*. One strategy for choosing strong promoters would be to screen the host's genome for sequences upstream from HEG. These sequence can either be lifted directly into the promoter cassette or further processed (dotted line) through promoter prediction software to design smaller core promoter sequences. Regulation and TF binding sites can be added separately by using online databases such as the Registry for Standard Biological Parts. Finally, restriction sites can be designed around the different promoter and regulatory elements to allow for rapid and straightforward interchange between different elements.

The choice between the different approaches would depend largely on their limitations. Optimization through random sequence screening would require a large

number of different constructs due to the statistical noise of many promoters that would be natively regulated, and sequences that would not carry promoter activity. On the other hand, a more focused approach combining the *in-silico* selection of promoter elements within genomic sequences and the introduction of known regulatory elements would depend on the accuracy of *in-silico* predictions.

5.6- Concluding Remarks

The use of the expression vector pQR1032, which contained the shorter staphylococcal promoter rplK, for the expression of CHMO resulted in an increase in the biocatalyst activity of the *E. coli* whole-cell system when compared with the previously established pQR239 expression system. This optimization was achieved through the constitutive transcription of the rplK promoter in *E. coli* , which negated the negative control of the Gram-negative pTac promoter, while allowing for the constitute expression of CHMO. The strong activity of the staphylococcal promoter in *E. coli* was a surprise, considering that the majority of its sequence was lifted directly from the Gram-positive *S. carnosus*. Using this optimized expression vector, pQR1034, we achieved maximum conversion rates of 0.55 g caprolactone L^{-1} h^{-1} while using *E.coli* TOP10 cultures in resting state conditions, with a specific whole-cell activity rate of 9.7 µmol caprolactone min^{-1} g DCW^{-1}. Total conversion of 20 mM of cyclohexanone was achieved after 6 hours of incubation. Performing whole-cell biocatalysis with actively growing cultures did not achieve total conversion of the ketone after 24 hours of incubation, and the caprolactone conversion rates were much lower than in resting-state biocatalysis. This is due to the fact that cultures used for the resting-state biocatalysis were at an higher concentration. The highest conversion rates were observed during the exponential phase, thus suggesting that the CHMO expression is at a maximum level during this period. On the other hand, the fact that total conversion was not achieved hints at the instability of the biocatalyst in the intracellular space during the stationary phase of bacterial growth.

The values from the resting-state biocatalysis are lower than activities calculated in other studies using pQR239 (Doig et al. 2003), and these difference might be due to changes in the experimental set-up that can be perceived from the total concentrations of cell cultures achieved by the end of incubation. In previous studies, *E. coli* concentrations of around 5 g DCW L^{-1} are achieved, whereas in our experimental protocols we obtain final cellular concentrations between 0.5 and 1 g DCW L^{-1}. A better comparison between pQR1034 and pQR239 can be observed through the NADPH oxidation studies, wherein culture lysates with pQR1034 exhibited 2 times higher NADPH oxidation rates than pQR239.

Unfortunately, the presence of the smaller staphylococcal promoter did not result in positive CHMO activity in *S. carnosus* cultures transformed with the expression vector pQR1034, despite the expression of the GenR in cultures cloned with pQR1032 and pQR1033. Correspondingly, neither whole-cell biocatalysis assays using resting or activity growing cultures generated any lactone products. No overexpression of CHMO could be detected on SDS-PAGE of supernatant and pellet fraction of growing cultures, hinting that perhaps the biocatalyst was not being expressed. However, similar protein gels done with *E. coli* cultures were also inconclusive in showing CHMO overexpression.

Cell wall permeability to cyclohexanone and caprolactone, as well as plasmid stability, were also assayed as factors contributing to the lack of CHMO activity. Substrate/product tolerance assays showed that *S. carnosus* TM300 had a higher tolerance to the ketone and lactone than *E. coli* TOP10, which could be attributed to a lower permeability of the former's cell wall to the compounds. However, the small difference in tolerance between the two strains does not account for the lack of biocatalytic activity. Indeed, if *S. carnosus* cell wall was impermeable to the substrate cyclohexanone and prevented its diffusion across the membrane, we would not observe ketone- mediated inhibition of cell growth. Plasmid stability assays showed that the vector pQR1034 is stable in *S. carnosus*, even after 3 days of incubation. Therefore, we also cannot attribute the lack of CHMO activity to vector instability.

The work we conducted on *S. carnosus* and the expression vector pQR1034 is only a superficial look on the characterization of the organism as a potential biocatalytic host or of the vector as a expression system for heterologous

biocatalyst expression. Unfortunately, due to time constraints, we were not able to extensively trouble-shoot the lack of pQR1034-mediated CHMO expression in *S. carnosus*. Indeed, factors such as the growth conditions of bacterial cultures could be crucial to the expression of active biocatalyst. We are also not certain if other metabolic processes within the bacterial strain outcompete CHMO for the ketone substrate. Using other ketone substrates in biocatalysis could test this hypothesis. Another factor that might dictate the rates and efficiency of CHMO activity is the codon usage bias of the protein. It is possible that optimization of the biocatalyst sequence to fit the codon bias of *S. carnosus* would improve its expression. All these enquiries will need to be considered in future work done on this project.

Despite the embryonic stage of the work done on *S. carnosus*, some of the results and trouble-shooting issues encountered throughout this project hint at the difficulties of this bacteria as a biocatalyst host. One of the potential problems is the physical robustness of the bacteria, which becomes undesirable when trying to extract active proteins from the intracellular environment. Directly related to this issue is the handling of genetic material, which as previous chapters highlighted, is cumbersome and requires methods that are either shear-intensive or require expensive chemical treatments. Finally, as the tolerance assays showed, the cell wall of *S. carnosus* did not generate a significant increase in tolerance to the biocatalytic substrates and products to justify its advantage in CHMO biocatalysis over *E. coli* strains. Of course, in order for us to give a definite answer regarding the suitability of *S. carnosus* as a biocatalytic host, we would need to test other biocatalytic systems. Also, there is also the possibility that pQR1034 could be further optimized for expression in the bacterial strain. However, the results from the biocatalytic experiments suggest that *S. carnosus* is not suitable for whole-cell CHMO biocatalysis.

Chapter 6: Conclusions

6.1- The Production of a Shuttle Vector for the Expression of Biocatalysts in *E. coli* and *S. carnosus*

One of the milestones of this research project was the production of a shuttle vector that allowed for the expression of heterologous proteins in both Gram-negative and Gram-positive bacteria. While the construction of this vector was initially motivated by the necessity to find solutions for problems generated by constructs used in previous studies, the ability to re-design a vector *de-novo* allowed us to expand on the designs that already existed for protein expression in *S. carnosus*. In fact, previous designs had focused on the integration of a couple of well-characterized promoters into a backbone that allowed for selection and replication in the staphylococcal strain, thus creating rigid expression systems that relied on the flexibility of these promoters for reproducible protein expression. The fact that the promoters had been extensively studied meant that these expression systems could be reliably used with ease in *in-vitro* and *in-vivo* conditions, thus reducing any need to change or search for better alternatives. Consequently, the constructs for protein expression in *S. carnosus* centred on the use of either the constitute lipase promoter, or the inducible xylose promoter for heterologous protein expression. While these promoters have been used for the expression of a myriad of different proteins, they had not been applied in the context of biocatalysis, where often a high level of protein expression is required.

Efficient whole-cell biocatalytic systems depend on the systematic optimization of the factors that directly or indirectly influence the performance of biocatalysis, and these include intracellular expression levels of biocatalyst. Subsequently, there is a constant pressure to engineer expression systems that

resemble those encountered in the biocatalytic host in order to minimize the negative effects of introducing novel or foreign metabolic circuits. Thus promoters are engineered to fit the requirements for optimal expression in the host organisms, while proteins sequences are re-shaped to reflect the codon bias of the latter. In this context, the lipase and xylose promoter systems are not optimized for biocatalyst production in *S. carnosus*. On one hand, these promoters are not native to *S. carnosus*, and while the genetic similarities between the latter an the staphylococcal species from which the promoters originate seem to hint at a common genetic mechanistic background, it is not a guarantee that these promoters are adapted for optimal expression in *S. carnosus*. On the other hand, the rigidity of the shuttle vectors containing the lipase and xylose promoters works as an antithesis to the adaptability required from engineered biocatalytic systems.

As a result, we attempted to address these potential limitations by designing a shuttle vector as a modular assembly, composed of functional elements that could be easily manipulated and replaced depending on the functional pressures of different biocatalytic systems. Essential to this design was a synthetic linker, which was designed as a reversible anchoring point for different staphylococcal promoters, thus allowing for the screening of the best promoter for biocatalyst expression. The addition of a flexible promoter region to allow for promoter screening and optimization is not a new idea, as the vector pCX15 (Wieland et al. 1995), used to characterize the xylose promoter in *S. carnosus*, was designed for a similar purpose. However, the design developed in this project expanded on this idea by adding the ability to optimize and screen promoter activity in both Gram-positive and Gram-negative bacteria. In addition, the modular nature of the design allowed for a fast adaptability to biocatalytic pressures not only in *S. carnosus* and *E. coli*, but also for other whole-cell systems by changing the replication and selection elements of the vector as required.

Figure 69 shows the different constructs produced from the initial vector design. The expression vector pQR1029 was created by joining the 3 main components of the vector design: the *E. coli* plasmid pTTQ18, a sub-section of the vector pCT20 containing the chloramphenicol resistance gene (CAM) and the staphylococcal origin of replication (oriS), and the synthetic promoter linker. The

construction of this vector was hindered by bottlenecks in the extraction of and manipulation of pCT20 from *S. carnosus*.

One of the problems was the fact that the thick peptidoglycan wall around *S. carnosus* rendered the latter extremely resilient to physical and chemical lysis. After trouble-shooting the methodologies that are routinely used to lyse Gram-positive bacteria, we were able to develop a simple protocol for DNA extraction that involved a single chemical lysis step with lysostaphin, followed by purification via a commercially available silica-gel membrane. This protocol was faster and more reliable than the alternatives, and we were subsequently able to use it routinely for plasmid and protein extractions with *S. carnosus*.

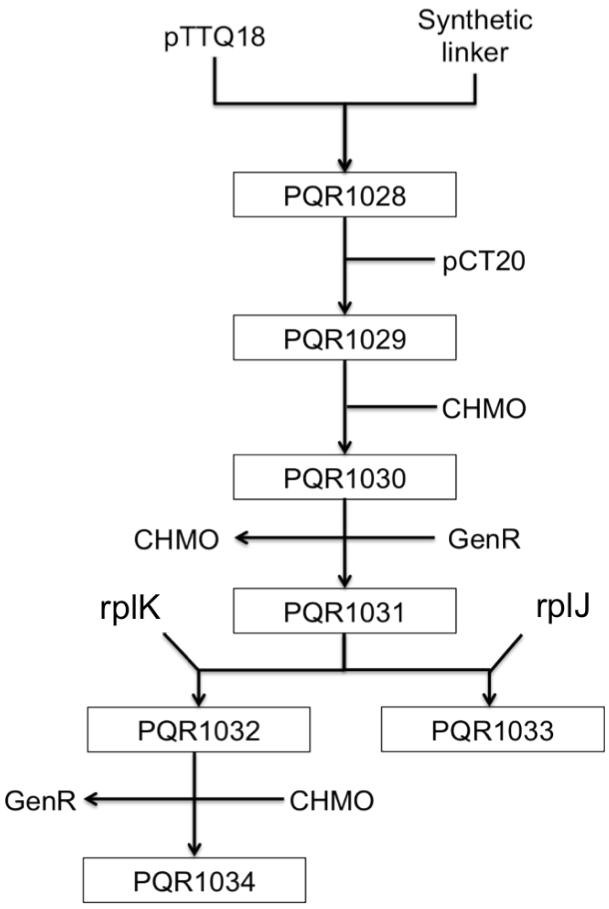


Figure 69- Diagram outlining the progression of the vectors constructed throughout this project. The arrows highlight the design strategy underlying the construction of each vector construct. For instance, the construction of pQR1029 revolved around the jointing of pQR1028 and the staphylococcal vector pCT20.

The other bottleneck in the restriction and ligation of pCT20 with pQR1028 related to a fault in the design itself. This ligation step relied on the restriction of both vectors with the endonuclease *Cla*I, which was inhibited by methylated sites on the target DNA. We did not account for this inhibition, and the *Cla*I site of the synthetic linker was designed with an overlapping methylation site, which subsequently prevented the restriction of pQR1028. This problem could only be solved by using a different *E. coli* cloning system that did not contain the methylation pathway.

The shuttle vector pQR1029 was transformed into both *E. coli* TOP10 and *S. carnosus* TM300, and shown to grant the adequate resistance phenotype to both species. Once the vector was proven to replicate and express the selective phenotype in both strains, the next step was to clone the CHMO gene into the MCS region of the vector, thus creating the third construct, pQR1030.

The CHMO expression profiles from pQR1030, as inferred from the NADPH oxidation assays, showed that the vector could be used to express the biocatalyst in *E. coli* through the activity of the pTac promoter from pTTQ18, but neither this or the staphylococcal promoter designed into the synthetic linker were active in *S. carnosus*. While we were expecting the pTac promoter not to show any activity in the latter, we were hopeful that the staphylococcal promoter, designated pRplK, which had been extracted from the staphylococcal genome as the region upstream of the ribosomal protein rplK would exhibit high promoter activity in the staphylococcal strain.

Exchanging CHMO for a gentamicin resistance gene (GenR) (pQR1031) challenged these results, by showing that pRplK had some, albeit low, level of activity in *S. carnosus*. Using the antibiotic resistance gene as a reporter, we were subsequently able to screen the genome of *S. carnosus* for stronger constitutive promoters. A shorter version of rplK promoter (pQR1032), in which both ends of the sequence were deleted, generated high levels of resistance to gentamicin in both *E. coli* and *S. carnosus*. This supported the assumption that genetic elements around the ends of the pRplK repress the expression of the rplK ribosomal protein, and are responsible for the lack of CHMO expression by pQR1030 in S*. carnosus*.

The staphylococcal genomic sequence upstream from the putative operon containing the ribosomal protein rplJ also generated a shorter promoter sequence

that could induce high levels of constitutive GenR expression (pQR1033), comparable to the shorter rplK promoter. Comparing the sequences for the two re-designed promoters highlighted sequence motifs around the promoters -35 and -10 RNA polymerase recognition sites that had been previously identified as conserved promoter elements within Gram-positive genomes: specifically, and A-rich area at position -42 from the transcription start, which was present in both promoters, and a TG motif just upstream from the -10 box, which was only present in the rplK promoter. In addition, the -10 box was mostly conserved between the two promoters, while the -35 box showed a greater degree of variation.

The pQR1032 vector, which inducing the stronger response to gentamicin, was subsequently used as an optimized version of pQR1029 for the expression of CHMO (pQR1034).

In the end, the different constructs produced throughout the project validated the principles behind the vector design strategy we used, by demonstrating the versatility of the core vector structure of pQR1029. In fact, through several simple modifications we were able to change the primary function of the vector from biocatalyst expression vector to reporter system for screening of promoter libraries. In addition, we also shown that this vector can be used as a powerful exploration tool that can be used to increase the general understanding of how genetic systems are regulated in organisms like *S. carnosus*.

6.2- The Promoter Region Upstream of the staphylococcal rplK Protein.

The results generated by the expression vector pQR1029 highlighted the role of the staphylococcal rplK promoter in the regulation of the expression of the 50S ribosomal protein L-10 in *S. carnosus*, which had not been previously reported.

In the gentamicin resistance assays (see chapter 4, section 4.11) the vector pQR1032, which contained a trimmed-down version of the rplK promoter region,

induced an estimated 7-fold stronger response to the antibiotic gentamicin than the precursor pQR1031. On their own, these results clearly indicate a difference in promoter strength that is not related to sequence differences in the core units of a promoter, but instead suggest the presence of areas within the region upstream of the rplK protein that exert a negative effect on the expression of this ribosomal protein. Indeed, since the shorter rplK promoter in pQR1032 is contained within the larger rplK promoter of pQR1031, we would expect both constructs to induce comparable responses to gentamicin in the absence of additional regulatory elements.

By complementing these results with the sequence analysis of the longer rplK promoter (Figure 70), we can start building a picture of the genetic elements surrounding the core promoter that could be responsible for protein down-regulation.

(i)- Catabolic repression

The area upstream of the -35 box contains a 36 bp stretch of palindromic DNA that could be involved in transcription regulation. A similar palindromic sequence had been previously identified upstream of the inducible xylose promoter (Wieland et al. 1995) characterized as a catabolite responsive element (cre) responsible for the glucose-mediated repression of the promoter (Figure 70).

Catabolite repression is a mechanism used by several bacteria and yeast to down-regulate the expression of specific catabolic genes in the presence of multiple carbon sources as a means to decrease the translation and transcription load and to maximize the conversion of specific catabolites (Hueck et al. 1994). In low G+C Gram-positive bacteria, catabolite repression works through the global regulator catabolite control protein A (CcpA). This protein is activated in the presence of glucose and other fast metabolized catabolites by association with the co-repressor Hpr, and binds to palindromic cis-acting sequences in promoter regions once activated, resulting on the up-regulation or down-regulation of transcription, depending on the location of the binding sites (Titgemeyer & Hillen 2002). Recently, a study of CcpA-mediated regulation in *S. aureus* (Seidl et al.

2009) demonstrated that CcpA acts upon several metabolic, transport, and virulence pathways of the cell. In addition, proteins involved in the xylose, sucrose, and lactose metabolic pathways were shown to be repressed by CcpA activation in *S. xylosus* (Jankovic & Brückner 2002). However, there is no evidence of a similar regulation occurring for proteins from the metabolic translation pathways in staphylococcal species.

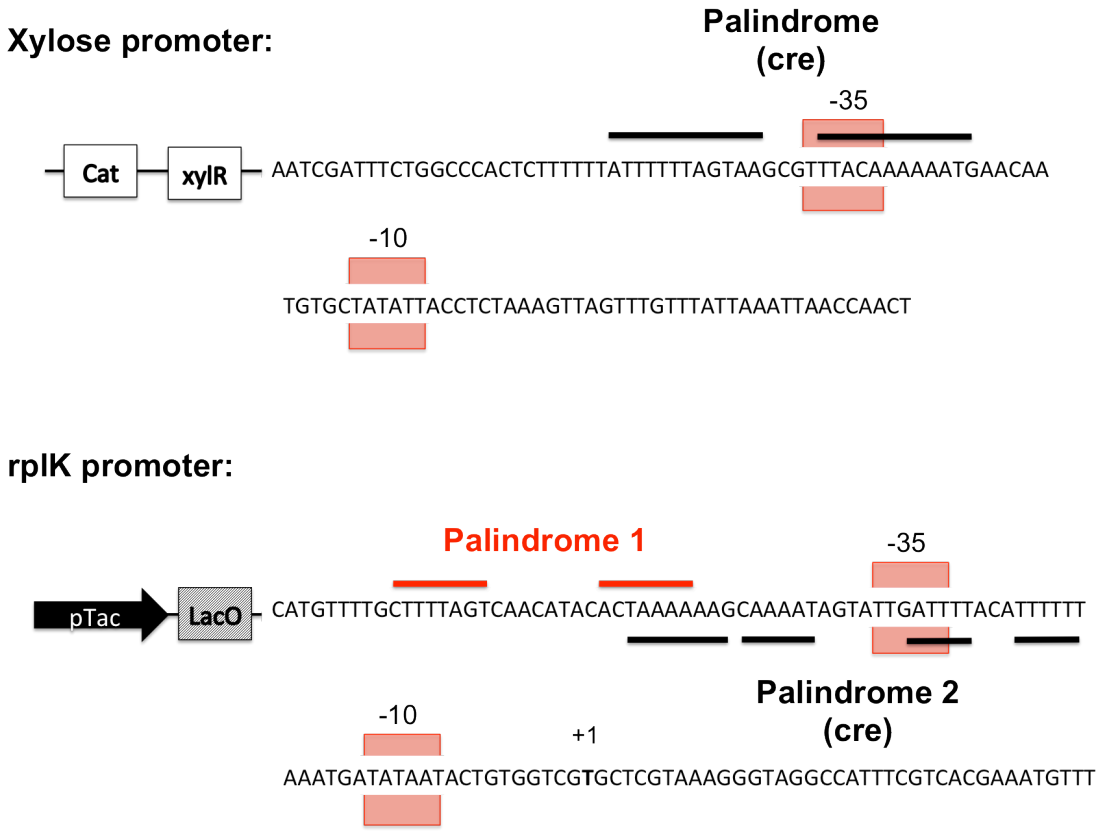


Figure 70- Graphical representation of the promoter regions from vectors pCX15 and pQR1030. The palindromic sequence previously identified by Wieland et al. (Wieland et al. 1995) as a CcpA binding site is highlighted by black horizontal lines. Black horizontal lines also outline a similar palindromic region in the rplK promoter of pQR1030, which we have defined as palindrome 2. In addition, another palindrome, highlighted by red lines, was also identified upstream, with one side of the palindrome overlapping palindrome 2.

It is thus surprising to see a cre-like sequence in the promoter region of the rplK protein, as it suggests that this ribosomal protein is also controlled by the presence or absence of specific catabolites. While there have been reports of ribosomal proteins being up-regulated in the presence of glucose in *E. coli* and

Bacillus subtilis (Gosset et al. 2004; Lorca et al. 2005) there hasn't been any conclusive evidence to establish a link between this effect and the CcpA regulation system. In addition, ribosomal proteins are essential components of the translation machinery, without which protein translation cannot occur, and thus it would be expected for these proteins to be maintained at a steady level of constitutive expression.

There are two possible hypotheses for the presence of a catabolite regulatory site upstream of the rplK promoter. The first hypothesis is that this site works in a positive manner, by allowing CcpA-mediated up-regulation of the ribosomal protein rplK, which fits with recent reports suggesting that CcpA binding sites upstream from the promoter elements are used as anchors that increase the affinity of RNA polymerase to the promoter (Titgemeyer & Hillen 2002), as opposed to CcpA sites within the promoters, that induce transcription repression (which is the case with the xylose promoter). However, this hypothesis contradicts the results we obtained from the gentamicin resistance assays, in which the presence of extra elements upstream and downstream from the promoter clearly had an inhibitory effect on protein expression levels.

Instead, the results seem to suggest that the cre sequence upstream of the promoter works as a repressor that induces down-regulation of rplK as levels of glucose increase. This hypothesis fits with the theory that bacteria and yeast down-regulate protein expression to maximize catabolite conversion. As the overall level of protein expression decreases, ribosomal proteins are also down regulated to avoid wasteful surplus. Therefore, it is possible that, in order for the glucose down-regulation signal cascade to be fast and effective, some ribosomal proteins are directly repressed by CcpA and become a limiting factor for ribosome assembly. The theory that rplK is directly repressed by high glucose levels also supports reports in which addition of glucose to growth media had an inhibitory effect on growth of *S. carnosus* (Tjener et al. 2004). Comparison between the expression levels of pQR1031 and pQR1032 in *S. carnosus* suggests that this repression is not maintained throughout the growth cycle of bacterial cultures, and could be directly related to the starvation of the carbon source in the media. Figure 71 shows the proposed mechanism for CcpA repression of the rplK promoter.

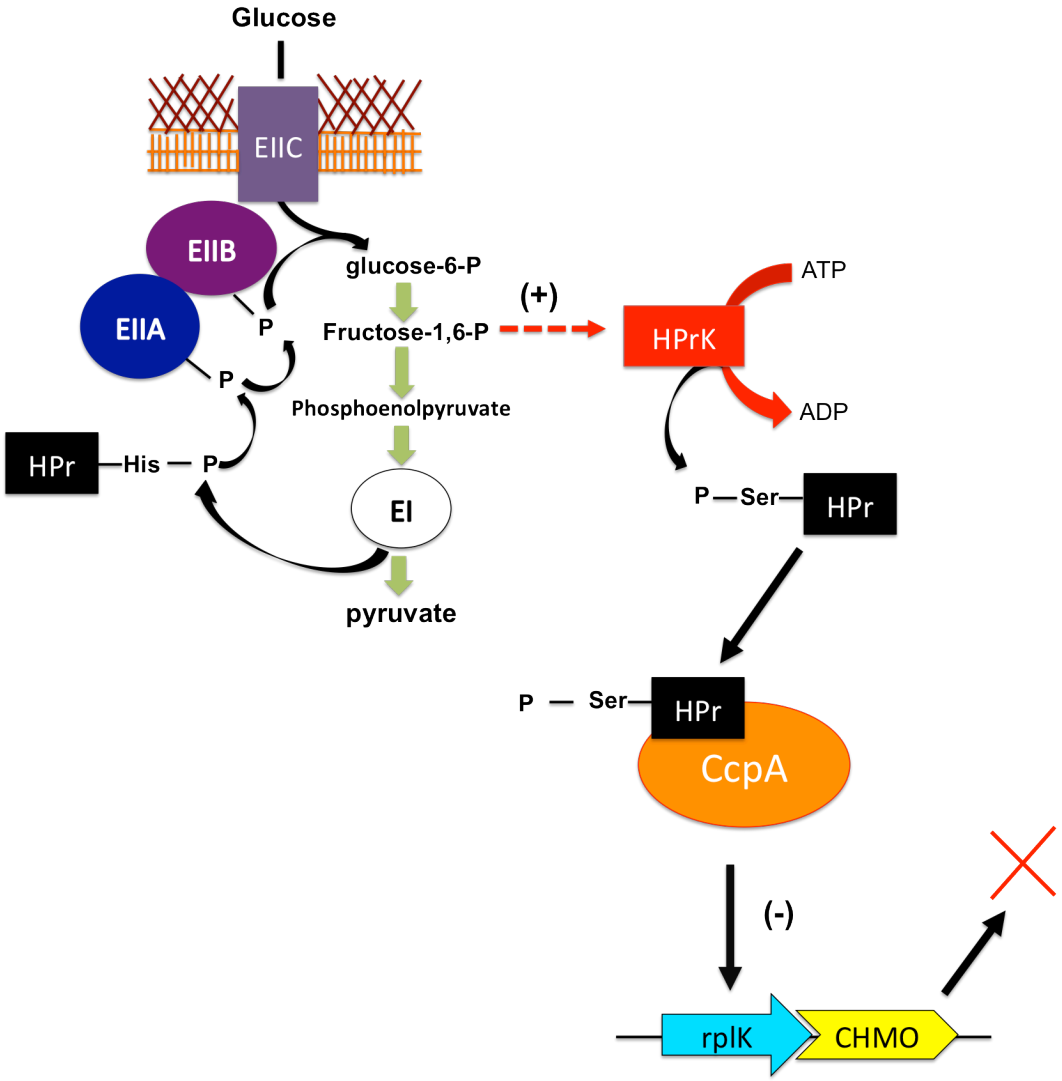


Figure 71- Diagram outlining the proposed mechanism by which the rplK promoter might be down-regulated by catabolic regulatory protein CcpA . The regulation depends on the phosphorylation state of a energy coupling protein of the phosphotransferase system (PTS), HPr. Depneding on the residue in which phosphorylation occurs, HPr can function as either a triggering factor for the transport of the carbon source by the EIIABC complex, or as an activator of CcpA. The two phosphorylation states complement each other as follows: The phosphorylation of a histidine residue in HPr is phosphoenolpyruvate (PEP)- dependent process that involves the pyruvate kinase EI. The histidine-phosphorylated HPr subsequently triggers the intake of the carbon source. In turn, the increasing levels of metabolites from glycolysis of the transported carbon source, such as fructose-1,6 biphosphate, stimulate another kinase, HPrK , which phosphorylates HPr in a serine residue. Seine-phosphorylated HPr will function as activator of CcpA. Thus, histidine phosphorylation of this protein triggers a positive-feedback loop that fuels self-regeneration through glycolysis, and induces activation of CcpA via HPrK.

By comparison, the native rplK promoter region exhibited very little or no activity in *E.coli* TOP10. Correspondingly, the tight regulation of pTac was not

negated by the insertion of the staphylococcal promoter, as can be seen from the NADPH oxidation assays with pQR1030 (see chapter 4, section 4.6). One reason for this outcome could be the fact that the transcription machinery of *E. coli* is not well adapted to operate with foreign Gram-positive promoter sequences and therefore could not induce expression using the rplK promoter. The genetic difference between the two systems might have created a barrier that would prevent any interchange of sequence functionality. However, by removing most of the rplK region to create a shorter promoter, we were able to not only increase the level of CHMO expressed in *E. coli* as compared with the pQR1030 and pQR239 constructs, but also bypass the inducible nature of pTac by generating constitutive expression of the biocatalyst.

Thus by showing that the smaller re-designed genomic rplK and rplJ staphylococcal promoters in pQR1032 and pQR1033 were active in *E. coli* TOP 10, these results suggest that there is a common genetic language that can be used by both the Gram-negative and Gram-positive bacteria to induce transcription. In addition, the results also indicate that rather than being inactive, elements within the rplK region in pQR1030 are actively preventing the *E. coli* mediated activation of the core rplK promoter, perhaps by employing a similar regulatory mechanism to that which was proposed for *S. carnosus*.

The catabolite regulatory pathway in *E. coli* works via the activation of the catabolite activator protein (CAP) (Titgemeyer & Hillen 2002). CAP exists as a dimer that is activated upon association to cyclic AMP, which in turn is converted from ATP by adenylate cyclase. Upon activation, the CAP dimer binds to a 22 bp palindromic sequence upstream from the promoter and induces transcription by association with the CTD domain of the RNA polymerase (Lawson et al. 2004). Alignment of this consensus sequence with the region upstream of the rplK promoter does not show significant homology between the two sequences. In addition, as an activator protein, CAP functions as a positive inducer of transcription, which does not account for the inhibition of CHMO expression in pQR1030. Since catabolite inhibition in *E. coli* involves a negative feedback loop in the activation of CAP, it is therefore improbable that the rplK promoter region is affected by homologous regulatory systems in *E. coli* and *S. carnosus*.

The involvement of Ccpa in the inhibition of the rplK promoter could be tested by incubation *S. carnosus* cultures containing the reported vector pQR1031 under different concentrations of glucose. If CcpA was indeed responsible for the repression of the rplK promoter, cultures grown with higher concentrations of glucose would be less tolerant to the antibiotic gentamicin as a result. Conversely, a different carbon source could be used to test if the repression signal is solely glucose sensitive. The present literature on *S. carnosus* is unclear as to what alternative carbon sources could be used for bacterial growth, and thus some experimentation would be require to access the best alternatives. However, since the genome of the staphylococcal strain contains Lac genes involved in lactose metabolism, we would hypothesize that this compound would work as a suitable replacement for glucose. Acetate could probably also be used due to the ability of staphylococcal bacteria to grow in high concentrations of the compound, with the caveat that pH of the extracellular environment would have to be tightly controlled to prevent expression variations resulting from media acidification.

(ii)- Unconventional DNA secondary structures

Another possible source of rplK-mediated inhibition in *E. coli* and *S. carnosus* may be the A and T-rich inverted repeats (IR) of the palindromic sequence upstream of the promoter core, which under physiological conditions can form cruciform secondary structures in double-stranded (ds) DNA, and hairpins in single-stranded (st) DNA and RNA (Figure 72). The mRNA generated from pTac transcription would include the whole rplK region, including the IR element, so it is possible for the latter to form a hairpin secondary structure that would inhibit translation of CHMO. However, this hypothesis was disregarded because the inhibition of CHMO expression was not observed at the translation level, as induction of the pTac promoter with IPTG still generated levels of CHMO activity that were comparable to the control pQR239.

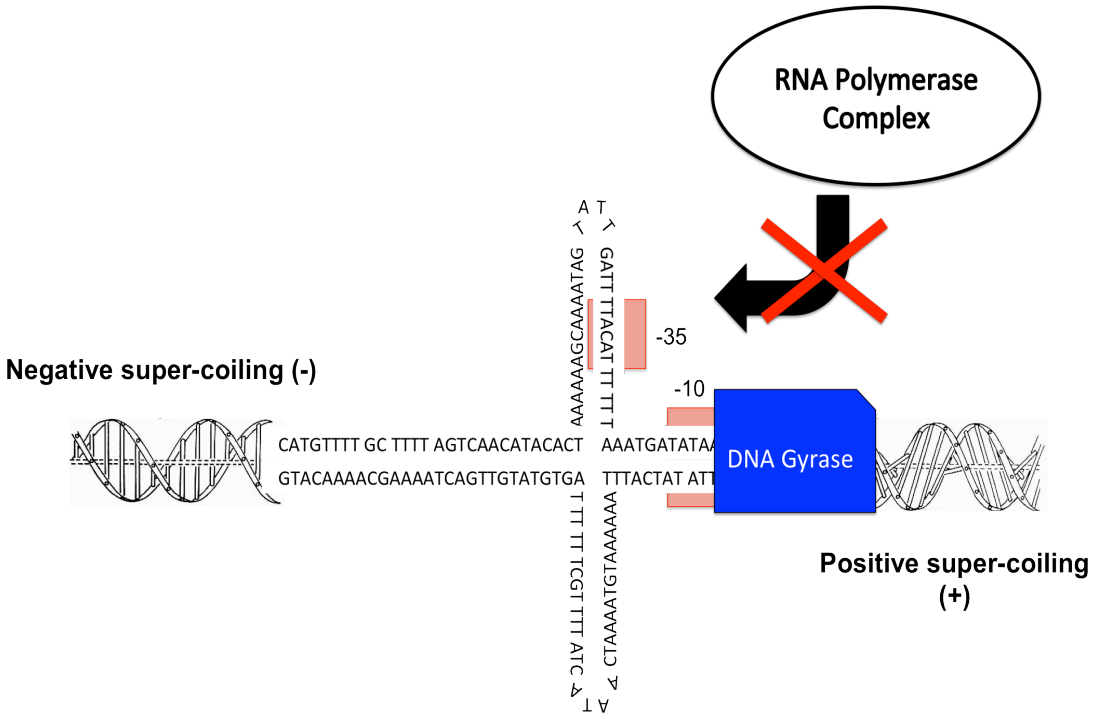


Figure 72- Graphical representation of dsDNA cruciform formation during the transcription process. As the DNA gyrase unwinds the area around the promoter region to allow for RNA Polymerase binding, it generates extra tension in the DNA sequence around this area: DNA downstream of the gyrase will be positively super-coiled , while the DNA upstream will be twisted in the opposite way through the longitudinal axis of the molecule (negative super-coilling). The relaxation around the promoter area will allow the palindromic region to for a cruciform structure that will prevent the binding of the RNA polymerase complex to the -35 element of the promoter.

Alternatively, cruciforms are DNA structures in which the two complementary IR elements form cis-annealing hairpins that extrude from the double-stranded DNA. These structures are not genetically stable and only occur through openings in the double-strand, which allows for intra-strand base pairing. Thus, cruciform formation is favored in events where negative supercoiling occurs, such as replication and transcription (Bikard et al. 2010). During transcription, DNA topoisomerases are employed to relieve the positive supercoiling occurring downstream of the transcription site by unwinding the ds DNA. As this unwinding occurs, the conformation of the DNA changes, so that areas downstream of the transcription machinery are positively supercoiled, while areas upstream are negatively supercoiled (Wang & Lynch 1993). There is evidence to suggest that IR

extrusion occurs in transition areas between the negative- and positive-supercoiled DNA in a transcription-dependent manner in *E. coli* (Dayn et al. 1992; Rahmouni & Wells 1992). In addition, this cruciform formation also seems to be dependent on the thermodynamic kinetics of IRs, with long (dA-T)n strands being more prone to form self-annealing hairpins structures in conditions that allowed for DNA negative supercoiling (Dayn et al. 1991) .

More important to our analysis is the association between DNA extrusion and transcription regulation. Several reports have highlighted that the presence of IR in the promoters regions led to promoter inhibition under physiological conditions, suggesting a mechanism wherein the negative supercoiling occurring at the start of transcription promotes the formation of cruciform structures, which in turn occlude the promoter recognition sites from binding to the RNA polymerase complex (Figure 72) (Horwitz 1989; Singh et al. 1995). Therefore, there is a possibility that a similar mechanism is occurring around the -35 region of the rplK promoter. The 36bp (dA-T)n IR sequence upstream of the promoter would be thermodynamically favorable to the formation of a cruciform structure, which would prevent the -35 box from binding to the RNA polymerase. This would also partly explain the fact that only rplK-mediated expression is affected, while the pTac promoter that lies upstream from IR region can still be induced in the presence of IPTG, as there is no indications that cruciform formation downstream of a promoter reduces transcription. On the other hand, if the IR sequence upstream from the rplK promoter was a binding site for regulatory proteins in *E. coli* , one would expect pTac-mediated expression to be precluded, as the presence of a regulatory protein would physically clash with the progression of the transcription complex. A similar process could also be occurring in *S. carnosus*.

(iii)- Transcription factor (TF) recognition sites

We also cannot rule out the possibility of negative interactions between other transcription factors of the bacterial hosts and the sequences around the staphylococcal promoter that could down-regulate transcription. In fact, screening of the pQR1030 rplK promoter with 81 consensus sequences for transcription

factors in *E. coli* K12 and *S. aureus* COL (taken from the RegPredict database) identified several small sequences upstream and downstream from the -35 and -10 RNA polymerase recognition sites that could acts as potential binding sites for transcription repressors (figure 73A).

One of these repressor is the EbgR protein, which belongs to the LacI/GalR family of transcription regulators, and acts by repressing the ebg operon (Hall et al. 1989). The ebg operon codes for a β-galactoside that is homologous to the LacZ but much more inefficient, not being able to promote bacterial growth on lactose in its wild-type form (Hall 2003). The Ebg repressor itself has very low sensitivity to lactose, and no known strong inducer for this protein is known. In the comparative analysis of the rplK promoter, the putative EbgR binding site was pin-pointed to a sequence several base-pairs downstream from the transcription start site, so it is unclear if binding of the repressor to this sequence would preclude transcription in the same manner as the LacI repressor, by inhibiting the formation of a stable transcription initiation complex. Regardless, it is plausible to assume that the presence of the repressor on the path of the RNA polymerase would generate enough conformational clashes to hinder transcription, and that the inactivation of EbgR by IPTG is required to allow for the strong activity observed in IPTG – induced cultures.

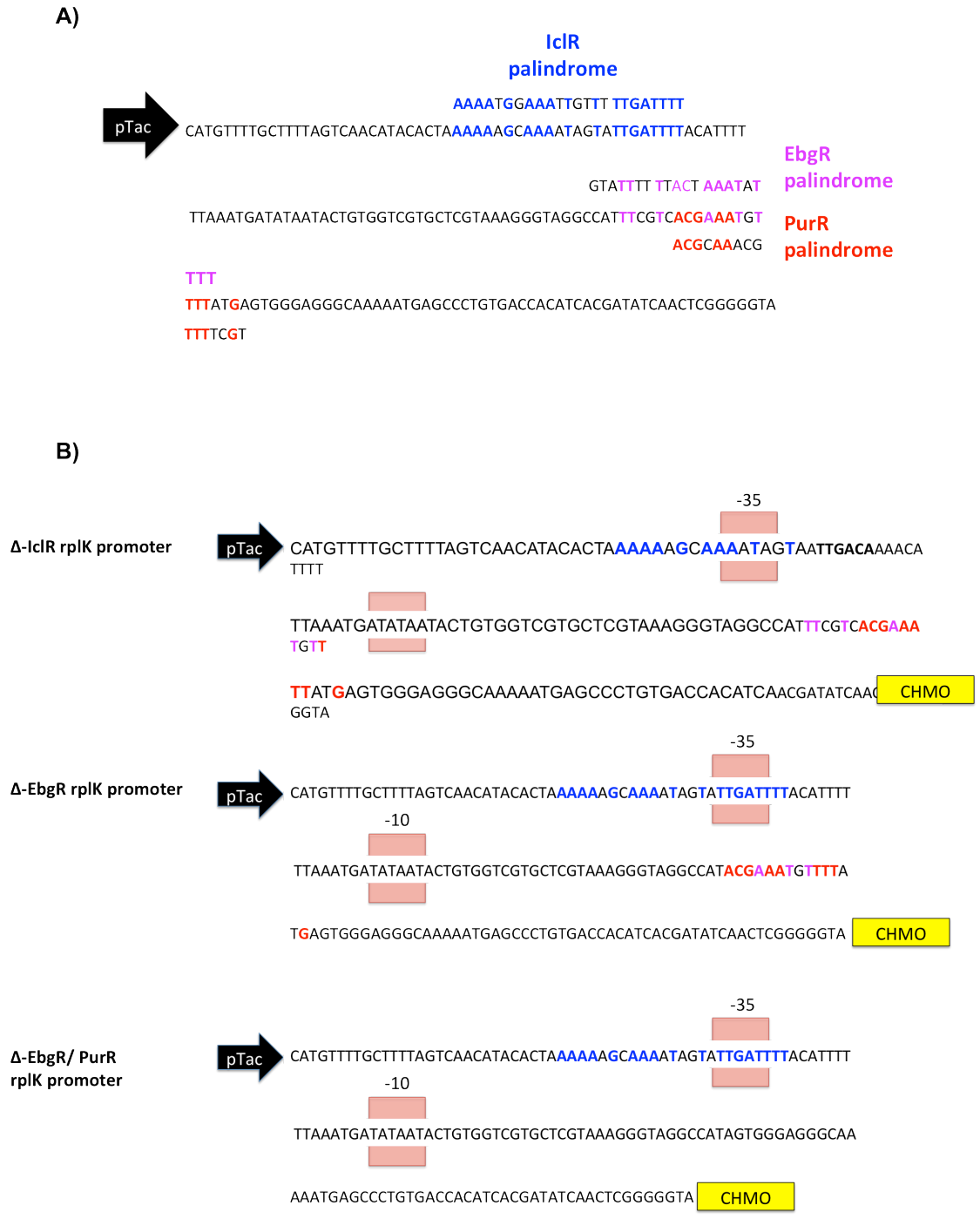


Figure 73- Homology between sites in the rplK promoter and TF binding sites for *E. coli* (A) and designs for truncated promoters that could be used to study the importance of these sites to transcription repression in *E. coli* strains (B). The homologies between the different TF sites and the promoter sequence are highlighted in different colours. Two of the sites, EbgB (purple) and PurR (red) , are situated downstream from the promoter region and overlap each other. The IclR site (blue) overlaps the -35 box, and subsequently the Δ-IclR construct is not a truncated version of rplK, but rather the -35 box of the promoter (TTGATT) is replaced for the consensus sequence TTGACA, which has a lower degree of homology to the rest of the palindromic sequence. By doing this, the IclR site is disrupted while avoiding the deletion of the core promoter.

Another palindromic sequence identified in the rplK promoter region that overlaps the putative EbgR binding site is homologos to the recognition site for the purine repressor PurR. This repressor actively suppresses several genes involved in pyrimidine and purine biosynthesis, as well as the conversion of adenosine monophosphate (AMP) (Anon 2011). Unlike LacI or EbgR, PurR requires the binding of a co-repressor, in this case guanine and hypoxanthine (Swint-Kruse & Matthews 2009). Strong repression was also shown to be elicited by an exogenous adenine stimulus (Cho et al. 2011). Since the co-repressors are compounds actively used in metabolic pathways, repression by PurR fluctuates between different stages of bacterial growth depending on the intracellular concentrations of these compounds, and as a result we would expect to observe intermittent strong CHMO expression via the rplK promoter throughout bacterial incubation. Since the levels of activity of non-induced pQR1030 culture are very low, it is unlikely that PurR would be the sole factor responsible for rplK repression.

The third transcription factor with a potential binding site in the rplK region is the IclR repressor. This factor actively prevents constitutive expression of proteins involved in acetate processing (Donald et al. 2001). In this case, acetate is the co-factor that binds to the repressor, thereby inactivating the repression of the aceBAK operon (Donald et al. 1996). As described previously, acetate becomes more prevalent during later stages of bacterial growth, when the oxygen becomes limiting and cells enter anaerobic fermentation in order to produce energy. Consequently, we would expect a spike of CHMO activity around the later stages of exponential and stationary growth if IclR was the sole factor responsible for rplK repression. The results from the biocatalytic assays suggest that this might not be the case, as biocatalytic activity still remained low after prolonged incubation.

The complexity of the feedback loops that regulate the different factor highlighted above makes it difficult to devise an experimental method to study each one individually. Both PurR and IclR use co-factors that are products of indirectly related metabolic processes. Correspondingly, it would be impossible to study the effects of one co-factor on the expression of CHMO without effecting the concentrations of the second. For instance, to study the possible binding of IclR, an experiment would be conducted in which *E. coli* cultures containing the vector

pQR1030 were grown in the presence of different acetate concentrations. However, the metabolic response generated by the presence of acetate in the extracellular medium would inevitably effect the concentrations of purines present in the intracellular medium. Consequently, it would not be possible to conclude if the biocatalytic activities observed under different acetate concentrations were a product of IcIR binding to the rplK promoter region, or if the guanine fluctuations in the cell during the metabolic response to acetate-mediated acidification of the extracellular medium also affected the binding of PurR to the promoter. Likewise, the similarities between the EbgR and LacI make it impossible to use external stimuli for the study of one factor without triggering the other. The only known anti-repressor for EbgR, lactose, would also have a similar effect on pTac expression.

Therefore, a more direct method of accessing the relevance of the putative transcription factor (TF) binding sites to the expression of the rplK promoter would involve the production of truncated version of the staphylococcal promoter DNA sequence so that the different binding elements are deleted sequentially. This would result in the construction of 3 variant of pQR1030, each containing a single TF binding site deletion, corresponding to each of the repressor proteins highlighted above. Change in CHMO expression in each of these constructs would directly correlate to the independent activities of the different repressors (Figure 73B).

The discovery of rplK repression in both bacterial strains used in the project is important for two reasons; firstly, it re-iterated the paradigm that bacterial promoters use a universal language that is valid between Gram-negative and Gram-positive bacteria; secondly, it highlighted the presence of genetic elements that can carry similar regulatory activities between different bacterial species. Discovering the nature of these activities would not only inform how the ribosomal protein subunit L-11 is regulated in *S. carnosus*, but would also highlight the parallelism between the genetic languages used for gene regulation in both *E. coli* and *S. carnosus*. In biotechnological terms, the discovery of regulatory elements that generate analogous responses in distinct bacterial species is very valuable because it informs the construction of universal promoters that can be used for heterologous protein expression cross-species.

6.3- The Promoter Screening Strategy

The choice of promoter for expression of CHMO in *S. carnosus* was primarily informed by the requirements of the biocatalysis systems, specifically that expression systems should be adapted to the cellular host to permit optimized biocatalyst expression. Therefore, instead of relying on promoters from related species that had been previously characterized, we decided to build a methodology for promoter selection that would allow for the use of promoters that are native to the biocatalytic host, and therefore offer optimized protein expression.

The initial approach to this problem was to search for regions upstream of proteins in the *S. carnosus* genome that were highly expressed and therefore could potentially be under the influence of a strong promoter. To identify these regions, we used the CAI analysis as a tool to predict highly expressed proteins, wherein the CAI values of a set of proteins were interpreted as estimates for the expression levels of that set. While the CAI analysis generated results that compared favorably with previous experimental studies on related species, employing this analysis on its own failed to produce a strong constitutive promoter from the genome of *S. carnosus*, as could be seen from the expression levels of pQR1030 in the staphylococcal strain (see chapter 4, section 4.7).

One limitation of the CAI analysis as a predictive tool for promoter strength is fact that CAI values are directly related to translation events. The analysis of the codon biases of different sequences is specifically related to the ease at which these sequences would be translated in organisms with strong codon selectivity. Thus, rather than accounting the level of transcription, CAI predictions are primarily focused on translation efficiencies of proteins.

This problem directly feeds into the second, and larger limitation of the CAI analysis. While attempting to give a broad overview of protein expression levels inside the cell, CAI uses a metric that is very heavily correlated to the specific cellular process of translation. However, protein expression is a multi-layered process that is regulated at various levels. For instance, the stability of mRNA is an important factor that is used as a means of controlling the gen population available for translation. The affinity of ribosomes to the mRNA is also tightly regulated, involving several factors and feedback loops that dictate how strongly

the ribosome binds and dissociates from the target sequence. As a result, the CAI analysis provides a very narrow window into gene expression that does not represent the totality of the factors that effect this process. It is therefore expected that the estimates from the CAI values do not represent the real levels of protein expression in cells.

Despite these limitations, we were still able to obtain valid estimates for the levels of expression of some of the proteins of the set we chose. Of the 46 proteins composing the query set, 39 generated CAI values that could be attributed to an high level of protein expression. Most of these are proteins concerned with housekeeping functions in the cell, such as translation or glycolysis, and therefore are expected to be present in high levels in cells at different stages of cellular growth. Proteins such as the ribosomal protein L7/L12 (rplL), which generated the highest CAI value in our analysis, had been previously shown to be one of the most abundant proteins in the cytoplasm of growing and stationary *S. aureus* cells (Kohler et al. 2005). Other translational oriented proteins such as the trigger factor Tig and the transcription initiation factor IF-2 (infB) that had previously been shown to be abundant in *S. aureus* also generated high CAI values.

Conversely, the glycolysis protein triosephosphate isomerase (tpiA) was estimated as HEG from the CAI analysis, while proteome studies showed that this protein was only present in high levels in growing *S. aureus*, but did not exhibit the same levels in stationary cells. This case re-iterates the limitations discussed previously, as the CAI analysis does not account for the regulation that inevitably occurs for glycolysis proteins. However, we did partly account for such limitations when selecting promoters. Thus proteins that were associated with catabolic networks, such as glycolytic proteins, or proteins that are directly implicated in responses to extracellular stimulus, as is the case of membrane transporters and chaperones, were not considered as potential candidates after the CAI analysis.

In the end, the correspondence between the CAI analysis and the proteomic evidence from related species gave us the confidence to use the results from the former as the basis for the promoter screening strategy.

The inability to extract a promoter from the genome that would generate high constitutive levels of expression in the first screening highlighted another limitation with the approach we used that is not directly relate to the software tools

applied in the screening. This limitation was the lack of knowledge about the systems that regulate gene expression in *S. carnosus*. As a result, by lifting whole sequences upstream of putative HEG as potential promoters, we were exposed to the risk of lifting not only promoter regions, but also other regulatory elements that would interact with the host in an inhibiting manner. Indeed, this was the case of the region upstream of the rplK gene, which as we discussed previously might contain regulatory elements involved in a glucose negative-feedback loop.

Only by employing promoter prediction software that filtered through the sequences upstream of putative HEG for specific promoter elements, were we able to generate short promoters that allowed for higher levels of constitutive expression. Therefore, by combining the predictive ability of the CAI analysis with the promoter screening software, we were able to optimize gene expression in *S. carnosus*. Thus the combination of these tools is a valid strategy for extracting strong promoters from the genome of the cell host for biocatalytic purposes.

6.4- The Expression Profiles of the Re-Designed rplK and rplJ Promoters

As described in the previous section, the two shorter promoters designed from the regions upstream from the rplK and rplJ genes were composed of fragments from the *S. carnosus* genome that had been screened for specific -35 and -10 RNA polymerase recognition sites. In addition, the promoters were designed with restriction site overhangs that allowed for cloning into the vector construct pQR1031, as well as RBS elements downstream from the transcription starting point that were also lifted from the area upstream of the ribosomal genes. The resulting ligations products of these promoters with the reporter vector pQR1031 generated gentamicin resistance profiles that were not only an improvement from the expression profile of the precursor pQR1031 in both *E. coli*

and *S. carnosus,* but also hinted at transcription mechanisms that might be interchangeable between *S. carnosus* and other Gram-positive bacteria.

The comparable activities of the smaller rplK and rplJ promoters suggest a universal structure for *S. carnosus* promoters, which is very similar to the consensus recognition sites for Gram-negative bacteria. In fact, the rplJ promoter contained an intact -35 box (TTGACA) and a partially intact -10 box (TATATT), whereas in the case of the shorter rplK promoter the opposite is observed. In addition, the distance between these elements is 17 bp, which agrees with that stipulated in Gram-negative promoters. However, the fact that pTac did not show any activity in *S. carnosus* indicates that additional upstream elements (UP) are required as recognition sites for the transcription machinery of the staphylococcal strain. The presence of UP elements in Gram-negative bacteria such as *E. coli* has been widely reported, but the affect they have on transcription depends on the affinity between the RNA polymerase complex and the main promoter -35 and -10 recognition sites (Rhodius et al. 2012) . By contrast, strong transcription in the Gram-positive *Bacillus subtilis* seems to be dependent upon the presence of A- and T- rich sequences upstream from the -35 promoter element that theoretically binds to the C-terminus of the RNA polymerase (Meijer & Salas 2004). Another UP element, a -16 TG motif, was also identified as crucial for strong expression in *Bacillus subtilis* (Voskuil et al. 1998). Our results suggest a similar requirement for transcription in *S. carnosus*, as both small promoters contained these UP elements. In fact, by doing a comparative alignment of the sequences upstream of the genes with the highest CAI values, we also observed that UP elements highlighted above are present in most of the sequences (Figure 74).

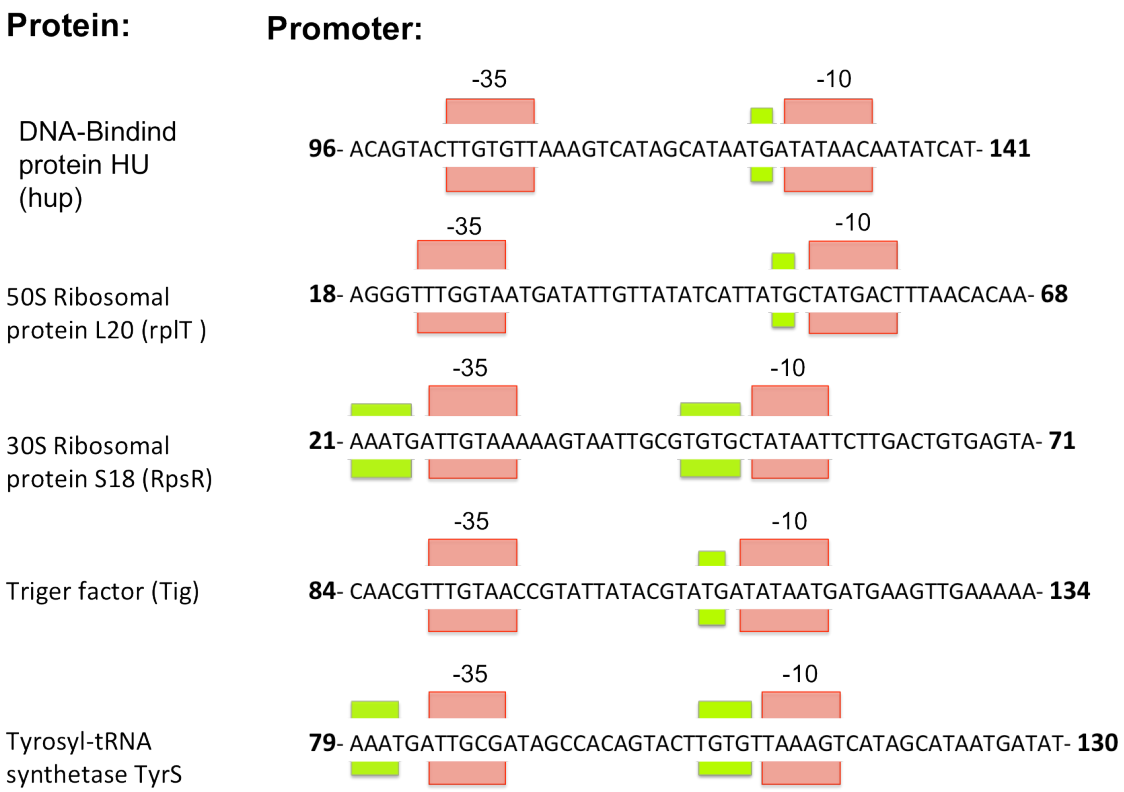


Figure 74- Sequences of putative promoter regions contained in the sequences upstream from 5 predicted highly expressed genes. The gene selection was based on the CAI analysis and the spacing between different genes in the genome. Thus, only sequences with a span longer than 40 base pairs between genes were considered. The promoter -35 and -10 elements are enclosed within the red boxes, while UP elements such as the -16 TG motif or the AT rich sequence upstream of the promoter are outlined by green boxes.

The difference in GenR expression between the two short promoters used in this project further elucidates the unequal relationship between the different UP elements and transcription strength. In fact, the UP element upstream of the -35 promoter element is more extended in rplJ, containing an extra 5 thymine residues. Despite this, a slightly higher resistance to gentamicin was observed in colonies containing the vector construct pQR1032. The fact that the rplK promoter contains an intact -16 TG motif, as well as an intact -10 TATAAT box, might account for the increased activity of the latter. On the other hand, the results clearly indicate that, while the presence -16 UP element resulted in an increase in promoter strength, it was not as essential to the overall expression as the UP element upstream of the -35 box, which was present in both promoters. Furthermore, the fact that extending the latter region did not produce a change in activity suggests the presence of 4

adenine residues immediately upstream from the -35 box is enough to induce strong transcription.

The fact that the two staphylococcal promoters allowed for constitutive expression of GenR in *E. coli* was surprising, albeit not all together unexpected. Since we used promoter prediction software tools that had been modeled on Gram-negative expression systems to filter through the *S. carnosus* genome, the probability of selecting sequences that could also be recognized by the *E. coli* transcription machinery was high. Nevertheless, the ability of *E.coli* to use promoters from the *S. carnosus* genome has not been previously reported.

In addition, the lack of protein expression from the longer staphylococcal promoter RplK in both *E. coli* and *S. carnosus* suggests a common negative regulatory system between these bacteria that interacts with regions upstream of the rplK promoter, resulting in promoter inhibition. Once these sequences were deleted from the core promoter, in the case of pQR032, both organisms were able to generate high levels of constitutive expression.

More importantly, the results with pQR1032 and pQR1033 clearly demonstrate how reverse engineering of genomic promoters from *S. carnosus* allowed for the creation of hybrid promoters that can be efficiently used in both Gram-negative and Gram-positive bacteria.

6.5- Expression of CHMO in *E. coli* TOP10

Previous studies of CHMO-mediated whole-cell biocatalysis have mainly focused on *E. coli* as host cells. In turn, the best characterized vector system to work in this bacterium for the expression of CHMO is the vector pQR239. Thus, we based the qualification of the performance of the vectors pQR1030 and pQR1034 on the direct comparison with pQR239, which was considered to be the standard for CHMO expression in *E. coli*.

However, the comparison studies done on the CHMO-mediated NADPH oxidation of lysates containing the different constructs presented some doubt about the nature of the pQR239 stocks used throughout this project. Specifically, the fact

that pQR239-mediated CHMO expression could be modulated by IPTG suggested that the plasmid was in fact different from that reported in the literature (O'Sullivan et al. 2001). CHMO is under the regulation of the pBAD promoter in pQR239, which is an L-arabinose promoter. This promoter is tightly regulated by the protein araC, which exerted a strong repression of protein expression in the absence of the activating substrate L-arabinose (Khlebnikov et al. 2002). In the presence of L-arabinose, the promoter is activated in a non-linear fashion, and therefore modulation of protein expression using this promoter is not possible. Despite this, the pBAD promoter was chosen for CHMO expression because it did not allow for leaky expression of the biocatalyst, and activation required the inexpensive substrate L-arabinose. Contrary to the results obtained in the NADPH oxidation assays, very low or no CHMO activity from pQR239 should be expected in *E. coli* cultures grown in the absence of L-arabinose. Additionally, IPTG should not induce an increase in pBAD expression. It is therefore possible that the pQR239 bacterial stocks contained instead the precursor vector pQR210, from which CHMO was lifted up to create pQR239. In this plasmid, CHMO is under the influence of the pTac promoter, which would explain the results from the NADPH oxidation assays. Unfortunately, we were not able to confirm this due to time constraints.

The possibility that pQR239 was in fact pQR210 offers a new perspective on the activity of the long rplK promoter in pQR1030. Since both vectors contain the same promoter and origin of replication (ColE1) for *E. coli*, we should expect the specific activity of cultures containing these vectors to be very similar under fully induced conditions. Consequently, the fact that pQR1030 exhibited higher specific activity hinted at the possibility that the longer rplK promoter is actually active in *E. coli*, providing for the extra CHMO expression. However, the assumption that pQR210 and pQR1030 have similar copy numbers inside cells and therefore express the same level of CHMO is not certain, since the vector pQR1030 also contains the staphylococcal origin of replication, which could contribute for the replication of the vector in E. coli.

Otherwise, the results from the NADPH oxidation assays corroborate the results from the gentamicin resistance assays with pQR1031, pQR1032 and pQR1033, and confirm that the smaller rplK promoter in pQR1034 allows for high levels of constitutive expression in *E. coli*.

They also showed that smaller staphylococcal promoter cancelled out the negative regulatory influence of the lacI repressor on pTac expression. This result is difficult to interpret because it suggests that the small rplK promoter is exclusively responsible for the high CHMO activity observed in pQR1034, and that the IPTG-mediated activation of pTac does not add to the overall expression of CHMO. In other words, the results suggest that the presence of second promoter upstream of a strong constitutive promoter does not add to the expression of downstream heterologous proteins. Further experiments would need to be conducted to confirm these results.

The whole-cell biocatalytic experiments done on resting and actively growing cells (see chapter 5, section 5.3) have a limited use for comparison with previously studies on CHMO whole-cell biocatalytic systems (Doo et al. 2010; Doig et al. 2003), due to the fact that the bacterial densities generated in these experiments were not comparable to the higher bacterial densities achieved in the previous studies. As a result, whole-cell activity of 9 μmol min^{-1} g DCW^{-1} calculated for resting state *E. coli* cultures containing the pQR1034 expression vector were much lower than activities of previously characterized strains (Doo et al. 2010; Doig et al. 2002).The disparity between activities might be a result from the sub-optimal conditions used.

However, the whole-cell biocatalytic experiments on actively growing cells were valuable to show that highest rate of CHMO activity occurs during the exponential phase. Additionally, the results show that CHMO activity significantly decreases during the stationary phase, which agrees with activity assays done on lysates after 24 hours of incubation. It is possible that, upon reaching stationary phase, constitutive expression of the biocatalyst is not viable, and the latter is either digested or forms inclusion bodies as a consequence of protein miss-folding. Alternatively, previous studies have hinted that the lower CHMO activity in actively growing cells might be a result of the inherent half-life of the protein (Walton & Stewart 2004), or due to competition for oxygen, which is also used for the general metabolic activity of the host and becomes a limiting factor during the later stages of bacterial growth (Baldwin & Woodley 2006). Another study conducted by Doig et al.(2003) suggested that the intracellular pH of bacterial cells could have an important role in the stability and activity of CHMO, which operates under an

optimum pH of 9. Correspondingly, decreasing the pH of a biocatalytic reaction from 9 to 7 resulted in a 5 times decrease in CHMO activity of cell lysates. The effects of pH would certainly partly account for the 10 times lower activity observed in resting cells (9 μmol min^{-1} $gDCW^{-1}$) when compared to the specific enzyme activity in lysates (88 μmol min^{-1} $gDCW^{-1}$).

6.6- Expression of CHMO inside *S. carnosus* TM300

No CHMO activity was observed in either the NADPH oxidation assays or whole-cell biocatalytic reactions performed with *S. carnosus* TM300 strains containing the expression vectors pQR1030 and pQR1034. Since the gentamicin resistance gene GenR had been expressed inside the staphylococcal strain using the same promoter constructs, the results led us to suspect that the lack of biocatalyst activity was a result of inherent limitations in the expression of the protein inside *S. carnosus*.

There are several factors that could contribute to this lack of protein expression. On the translation level, successful CHMO translation could be hampered by differences in codon biases between the protein and the host organism. Previous studies have shown the importance of codon biases as modulators of heterologous protein expression (Gustafsson et al. 2004; Gvritishvili et al. 2010).

CHMO is native to *Acinetobacter sp.* NCIMB 9871, a Gram-negative bacterium with a distinct genetic make-up from *S. carnosus*. Despite that, CHMO has been shown to be actively expressed in the Gram-positive bacterium *Corynebacterium glutamicum* (Doo et al. 2010), suggesting that the Gram-negative/Gram-positive codon bias disparity is not an important factor in the translation of the biocatalyst. However, *C. glutamicum* is in many respects more suited to express Gram-negative proteins than *S. carnosus.* Firstly, Gram-negative expression systems such as the pTac promoter can be used in *C. glutamicum*, thus removing the need for a dedicated promoter system for protein expression.

Additionally, looking at the codon preferences of C*. glutamicum* and *S. carnosus*, we observed that the former matches more closely to the codon bias of CHMO (Figure 75). Accordingly, the protein sequence contains several high frequency codons that are not prevalent in *S. carnosus*. For instance, the codon GCC, which codes for 12 of the 46 alanine amino acids in CHMO, is used in less than 10% of synonymous sequences coding for alanine in *S. carnosus*, and therefore could be considered to be a rare codon (Gustafsson et al. 2004). Other examples of *S. carnosus* rare codons that are significantly present in the CHMO sequence include CUG, which codes for 6 leucine amino acids in the protein, UCC, which codes for 3 serine amino acids, and ACC, used 12 times to code for threonine. Together, these correspond to 6 % of the protein being coded by codons with a very low prevalence in *S. carnosus*. By contrast, only 2.7% of CHMO is coded by rare codons in *C. glutamicum*. Although the percentage of rare codons used in the translation of the biocatalyst is not high, it is possible that it could represent a burden to successful translation, specifically If the rare codons are clustered together.

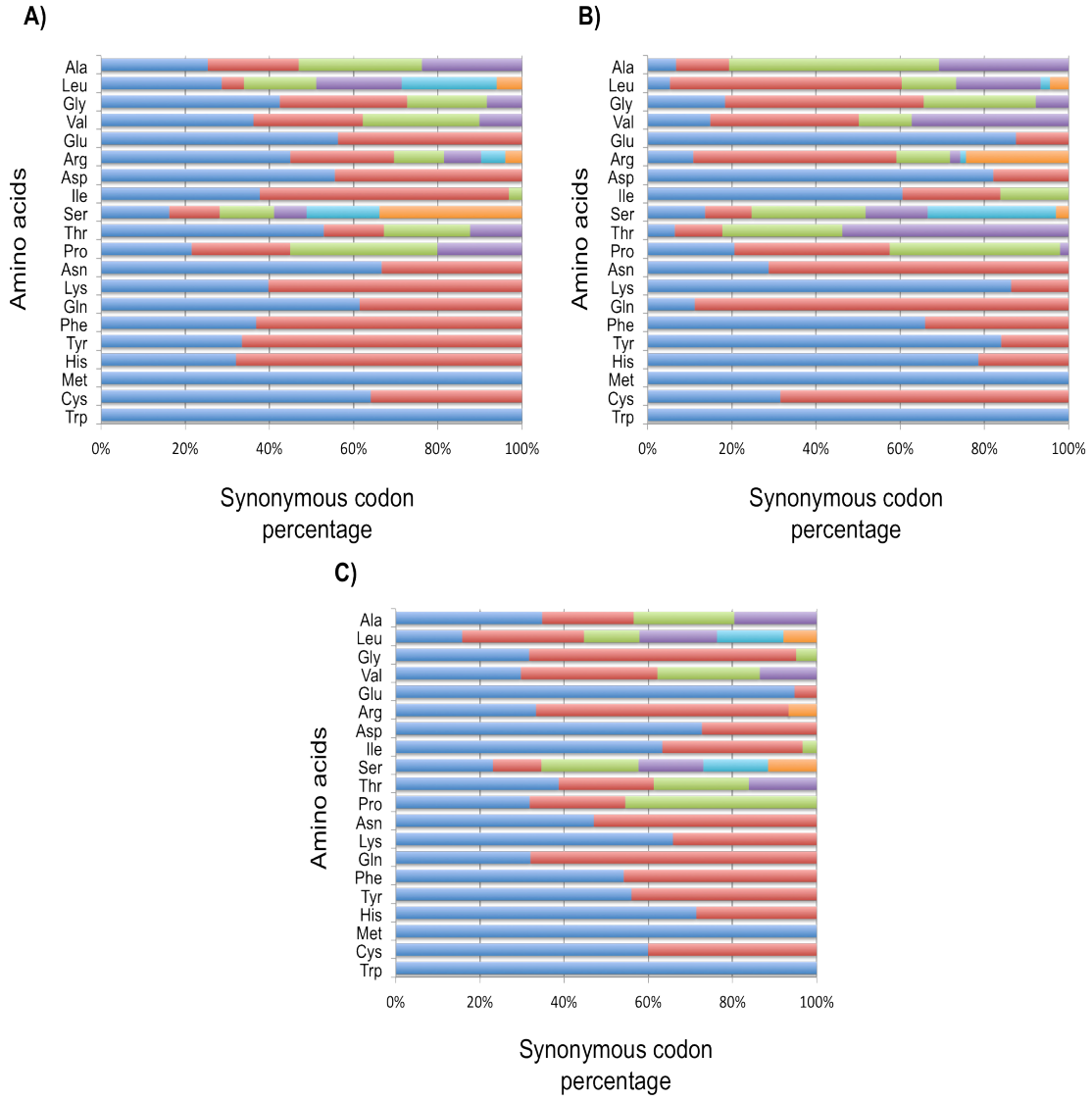


Figure 75- Codon bias barcharts for *C. glutamicum* (A), *S. carnosus* (B), and the protein CHMO (C). The different coloured bars correspond to different synonymous codons for a specific amino acid, and are displayed as a percentage of the total number of codons for that amino acid. The colours are consistent throughout the three charts, i.e. they correspond to the same codons in the three plots.

Post-translational events could also contribute to the inability of *S. carnosus* to produce active biocatalyst. Intracellular proteases that degrade mis-folding and aggregating proteins could play a role in the possible degradation of CHMO. One of these proteases is the ATP-dependent ClpP protease, which is highly conserved in Gram-positive genomes (Frees et al. 2007). This protease is an oligomer that depends on the binding of Clp ATPases to specific targets that are subsequently carried into the ClpP complex and digested. Different ATPases are normally

expressed at distinct points in the bacterial life cycle and recognize protein targets that are involved in regulatory and stress-response networks (Frees et al. 2007). More recently, the CLpP protease complex has been associated with the degradation of non-native proteins. Proteome studies with *Bacillus subtilis* have shown that overexpression of heterologous proteins triggered an increase in ClpP expression (Jürgen et al. 2001), while deletion of the protease gene in *Lactococcus lactis* resulted in an increase in expression of active heterologous protein (Frees et al. 2001; Loir et al. 2005). ClpP has also been shown to have a similar function in *S. aureus*, suggesting that it could actively hindering protein expression in the genetically-related *S. carnosus* (Michel et al. 2006; Frees et al. 2003).

The current proposed mechanism for ClpP-mediated degradation of non-native proteins involves the ATPase ClpC, which is coded by the ctsR repressor operon. Upon protein denaturing conditions, ClpC is titrated from a regulatory complex to bind to specific mis-folded or non-native proteins, and subsequently binds to the ClpP core oligomer to promote total unfolding and denaturation of the target (Frees et al. 2007). The uncoupling of the ClpC from the regulatory complex also triggers the binding of the complex to the repressor CtsR, which auto-regulates the ctsR operon. Thus ClpC expression works as positive feedback loop that is triggered by the presence of miss-folded proteins in the intracellular space. By analyzing the genome of *S. carnosus*, we were able to identify a putative CplP gene with high homology to the *L. lactis* and *B. subtilis* counterparts (Figure 76A). In addition, a CplC gene homolog was identified in an operon-like genomic segment that also contained the CtsR gene (Figure 76B). Thus, it is possibly that the CplCP complex acts in *S. carnosus* to prevent the accumulation of non-native proteins, and could consequently be responsible for the denaturation of CHMO.

Another ATP-dependent protease that could contribute to the degradation of overexpressed proteins is the membrane bound FtsH, a metaloprotease that is highly conserved through the bacterial genera and recognizes and digests unusual or unstable protein conformations (Ayuso-Tejedor et al. 2010; Li et al. 2004). A homolog of this enzyme can be found in the genome of *S. carnosus*, so we cannot rule out its possible role in the degradation of miss-folded CHMO.

A)

```
Staphylococcus carnosus  1    MN-LIPTVIETTNRGERAYDIYSRLLKDRIIMLGSAIDDNVANSIVSQLLFLQAQDAEKDIYLYINSPGGSVTAGFAIYD  79
Lactococcus lactis       1    MGyLVPTVIEQSSRGERAYDIYSRLLKDRIIMLTGPVEDGMANSIIAQLLFLDAQDNTKDIYLYVNTPGGSVSAGLAIVD  80
Bacillus subtilis        1    MN-LIPTVIEQTNRGERAYDIYSRLLKDRIIMLGSAIDDNVANSIVSQLLFLAAEDPEKEISLYINSPGGSITAGMAIYD  79

Staphylococcus carnosus  80   TIQHIKPDVQTICIGMAASMGSFLLAAGAKGKRFALPNAEVMIHQPLGGAQG--QATEIEIAATHILKTRAKLNKILAER  157
Lactococcus lactis       81   TMNFIKSDVQTIVMGMAASMGTIIASSGTKGKRFMLPNAEYLIHQPMGGTGQgtQQTDMAIVAEQLLKTRKRLEQILADN  160
Bacillus subtilis        80   TMQFIKPKVSTICIGMAASMGAFLLAAGEKGKRYALPNSEVMIHQPLGGAQG--QATEIEIAAKRILLLRDKLNKVLAER  157

Staphylococcus carnosus  158  TGQSIEQIEKDTDRDNFLTADEAKEYGLIDEVM---QPEK  194
Lactococcus lactis       161  SNRSLEQIHKDAERDHWMDAKETLEYGFIDEIMEN-NSLK  199
Bacillus subtilis        158  TGQPLEVIERDTDRDNFKSAEEALEYGLIDKILTHtEDKK  197
```

B)

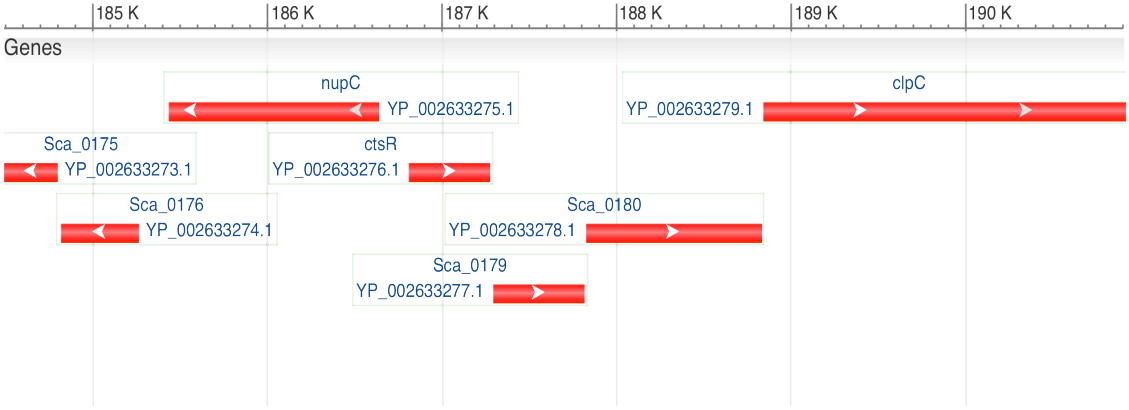


Figure 76- Analysis of the ClpP and ClpC protease subunits in the genome of *S. carnosus*. A) Protein sequence alignment between the ClpP of *S. carnosus* and counterparts found in *L. lactis* and *B. subtilis* . B) Graphical representation of the operon-like region in *S. carnosus* genome containing the ClpC ATPase. The repressor CtsR can be found upstream from ClpC, at about 187 K base pairs.

Finally, the presence of competing metabolic pathways that use the substrate or product of CHMO-mediated biocatalysis is also a possible factor in the lack of conversion observed in whole-cell experiments. Side reactions have been previously reported to be problematic for CHMO biocatalysis in bakers yeast, a problem that led to the shift to *E. coli* as a host for CHMO biocatalysis (Stewart et al. 1996). Analysis of the *S. carnosus* genome did not generate any positive hits for cyclohexanone dehydrogenases or lactone hydrolases. A 3-oxoacyl-ACP

reductase with 46% homology to the cyclohexanol dehydrogenase from Acinetobacter sp. NCIMB9871 was identified, but it is unlikely that this enzyme would compete with CHMO for cyclohexanone or caprolactone. On the other hand, the reductase uses $NADP^+$ as an acceptor, which could interfere with the oxidation rates of the monooxygenase. However, since *E. coli* strains also contain this enzyme, it is unlikely that it would contribute to the negative results.

As stated above, the data obtained in this project is insufficient to answer these queries, and further experimentation is needed to trouble-shoot the lack of biocatalyst activity.

6.7- *S. carnosus* as an Alternative Host for CHMO Biocatalysis

Whole-cell biocatalytic processes are complex systems involving several factors that are independent from the biocatalyst or biocatalytic reaction. Therefore, as this project exemplified, the choice of host is not straightforward and often involves the consideration of multiple criteria. While these criteria are heavily dependent on the choice of biocatalysis and biocatalytic reaction, some are general enough to be valid for most whole-cell systems. For instance, the ability of the host to stably express the biocatalyst intracellularly, and the genetic expression systems that are available for use in the host are important considerations to make, regardless of the biocatalyst. Likewise, the growth profile and optimum growth conditions of the host are factors that will directly affect the performance of the whole-cell process independently from the biocatalyst used. Other criteria, such as the permeability of the cell wall to substrates and products, or the presence of competing metabolic networks, are criteria that will be more or less impactful with regards to the choice of biocatalyst. For instance, when employing biocatalytic reactions where toxic substrates and/or products are used, the permeability and resistance of cells to exposure to these compounds becomes a major consideration for the choice of host.

From the results obtained in this project, we are able to make some conclusions about the fitness of *S. carnosus* TM300 as a host for CHMO biocatalysis (Figure 77). Growth studies showed that under similar conditions, *S. carnosus* reached faster growth rates and total growth than *E. coli* TOP10. This characteristic would be advantageous for whole-cell biocatalysis, as either biomass densities would result in increased total yields of the biocatalytic reaction. In addition, *S. carnosus* was not only shown to grow efficiently in conditions used for the Gram-negative counterpart, but it was also adaptable to more aggressive conditions that severely impaired *E. coli* growth. In fact, optimum *S. carnosus* growth was observed in media containing 2.5 % (w/v) sodium chloride (NaCl), which would be naturally selective against other bacteria species. The fact that high-salt, low pH media are already used in food industry to hinder contamination by pathogenic bacterial demonstrates the advantage of using *S. carnosus* as bacteria capable of growing in harsh environments.

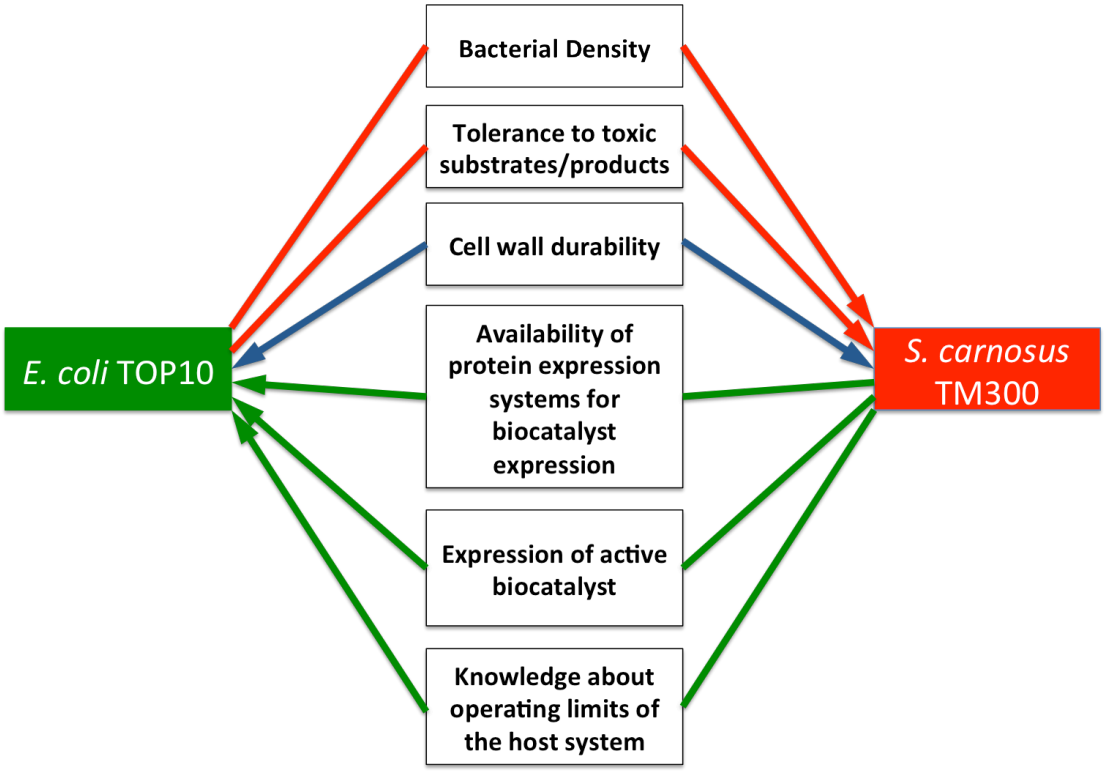


Figure 77- Graphical representation of factors that affect whole-cell biocatalysis and their relation with the two bacterial strains used in this project. The direction of the arrows and colours of the lines highlight the advantage of the strains in relation to the factor. The red directional lines indicate that *S. carnosus* TM300 is more advantageous as a biocatalytic host when considering bacterial densities and tolerance to toxic metabolites. By contrast, the green arrows indicate factors that favor the use of *E. coli TOP10* as a host system. The blue bi-directional blue arrows indicate that neither organism has the upper hand when considering the strength of the cell wall, i.e. there are advantages and disadvantages to using either.

Directly related to this is the shear-resistant nature of the Gram-positive bacterium, observed throughout the bacterial lysis protocols conducted in this project. This characteristic could be advantageous in retaining the stability and activity of intracellular biocatalyst in shear-intensive reaction conditions, such as when high levels of mixing are involved. However, the hardiness of the staphylococcal cell wall also works against the system, as it has been shown in several experiments where cell lysis was required. The harsh shear conditions used in this experiments were often inefficient without the pre-treatment with expensive enzymatic cocktails, and risked the stability of the biocatalyst due to over-heating or enzymatic interference. Additionally, the thick peptidoglycan wall greatly hindered attempts to clone and manipulate genetic material in *S. carnosus*.

In fact, all the protocols used routinely for *transformation* of *E. coli* were ineffectual for cloning into *S. carnosus*, and the protocols that resulted in positive clones had much lower yields. DNA extraction also required protocols that could potentially harm the integrity of the genetic material and made it difficult and cumbersome to analyze the plasmid content for the staphylococcus strain.

Other factors were also undesirable for biocatalysis. The lack of general knowledge about the staphylococcal host and the absence of a wide range of well-characterized genetic systems for the expression of heterologous proteins were responsible for the extensive amount of time devoted in this project to the construction of an efficient expression vector. On one hand we see this lack of knowledge as an advantage rather than a disadvantage in scientific terms, as our study of promoters from the staphylococcal genome has contributed to the general understanding of what promoter systems can be used for heterologous protein expression in this bacterium. However, the time investment and the cost that such studies demand is not economically viable to projects that rely on fast creation of highly efficient biocatalytic systems. This is one of the reasons why most of the current efforts regarding biotechnological advances in whole cell biocatalysis have focused on the optimization of already well-established systems, as opposed to searching for novel bacterial hosts. And while there is an argument for the latter approach as a means to bring biotechnological novelty to whole-cell biocatalysis, the lack of understanding about the operating rules and limits of novel cell systems can still be seen as a disadvantage, particularly when the benefits from using these systems are not inherently apparent. Such a case could be made of *S. carnosus*.

Even considering the higher tolerance to cyclohexanone and ε-caprolactone, together with the higher bacterial densities that would result in higher biocatalytic rates, the inability to express active CHMO in *S. carnosus* makes it an undesirable host for CHMO biocatalysis. In fact, the results from the biocatalytic assays suggested that intracellular expression of non-native proteins in *S. carnosus* is not as straightforward as in more conventional systems, and requires careful consideration of factors such as the codon bias of the protein gene and the proteolytic degradation of overexpressed proteins.

In conclusion, the advantages of using *S. carnosus* as a host for CHMO biocatalysis do not seem to out weight the disadvantages of working with a poorly

characterized system that is difficult to clone and manipulate, particularly when no biocatalyst expression was observed. However, the results from this study are only preliminary, and a more in-depth trouble-shooting process would have to be conducted to give definite conclusions about the expression of CHMO in *S. carnosus*.

6.8- Future Perspectives

6.8.1- The Use of Alternative Biocatalysts

The PhD project described in this manuscript focused on CHMO as the biocatalyst. And while the results from this study heavily infer that *S. carnosus* might not be suitable for whole-cell biocatalysis with this protein, the study of a single biocatalytic system is insufficient to conclude that the staphylococcal strain is inadequate for whole-cell biocatalysis in general. Correspondingly, to answer such a query would require the expression and study of the performance of several other biocatalysts in *S. carnosus*.

CHMO is the most widely studied Baeyer-Villiger monooxygenase, but it is not the only one to be cloned and expressed in bacterial systems. Five other monooxygenases have been studied in the context of recombinant bacterial systems. Two of these, the steroid monooxygenase (STMO) and the cyclododecanone monooxygenase (CDMO) were encountered in Gram-positive *Rhodococcus* species and subsequently cloned into *E. coli* (Kostichka et al. 2001: Morii et al. 1999).

CDMO, first cloned from *Rhodococcus ruber* CD4 and SC1, was shown to be able to efficiently convert long chain cyclic ketones in recombinant *E. coli.* This monooxygenase is an interesting and potentially valuable candidate biocatalyst, as it not only catalyses the degradation of compounds contained in petroleum (Kamerbeek et al. 2003), but also produces lauryl lactone, which functions as a pharmaceutical building block and is not currently chemically produced (Yang et al. 2009). Additionally, the bulky aliphatic nature of substrates for CDMO represents a

big shift in size from the single and dual ring substrates of CHMO, and therefore would allow us to study the permeability of *S. carnosus* to different substrate sizes.

Two variants of the STMO that performed conversion of steroid compounds were first identified in *Rhodococcus rhodochrous* and Cylindrocarpon radicicola, but only the formers' monooxygenase was implemented in a whole-cell biocatalytic system with *E. coli* as the host (Kamerbeek et al. 2003). *E. coli* BL21 (DE3) cultures expressing the STMO were able to achieve protein yields about 40 times greater than STMO expression in its native strain (Morii et al. 1999). However, no further biocatalytic studies were performed on whole-cell conversion of steroids such as progesterone. Thus, on one hand the study of STMO in *S. carnosus* as a biocatalytic system would be very valuable in the categorization of the enzyme as a biocatalyst. On the other hand, the lack of information about this protein does not fulfill the purpose of comparing *S. carnosus* as a biocatalyst host to other established whole-cell systems. It is worth noting that although STMO and CDMO are native from Gram-positive Rhodococci, these strains are genetically distinct from *S. carnosus*, with GC-rich genomes, as opposed to the AT-rich staphylococcal counterpart. Correspondingly, the both monooxygenases code 12 to 17% of their sequences with *S. carnosus* rare codons, which is twice as much as the CHMO sequence. As suggested above, this disparity between protein codon biases and host codon preferences might create bottlenecks for the subsequent expression in the host, and correspondingly codon optimization of both protein sequences would have to be performed before expressing them *in S. carnosus*.

Another Baeyer-Villiger monooxygenase, cyclopentanone monooxygenase (CPMO) that was isolated from *Comamonas* sp. NCIMB 9872, has been shown to share an substrate specificity overlap with CHMO, although with different enantio-selectivities (Kamerbeek et al. 2003). Since this enzyme has mostly been studied in the context of whole-cell systems (Iwaki et al. 2002; Bes et al. 1996), there is enough data from which to compare the performance of *S. carnosus* as a biocatalytic host for CPMO biocatalysis. Another aspect also makes this monooxygenase an interesting candidate. There is a high degree of disparity between the codon bias of the CPMO gene and that of the staphylococcal host, with 43% of the sequence being composed of rare codons in *S. carnosus*. This

difference could be exploited as an alternative way to access the impact of codon bias to heterologous protein expression in *S. carnosus*.

Finally, the 4-hydroxyacetophenone monooxygenase (HAPMO) was initially encountered in *Pseudomonas fluorescens* ACB growing on 4-hydroxyacetophenone. This enzyme was characterized as the first Baeyer-Villiger monooxygenase able to catalyse aromatic ketones, as well as a broad range of aliphatic compounds (Kamerbeek et al. 2003). Due to the fact that its native host contains enzymes that degrade the products from HAPMO oxygenation, this enzyme has been cloned and overexpressed in *E. coli* under a T7 promoter system (Kamerbeek et al. 2001). Another study was able to demonstrate the use of this monooxygenase as a biocatalyst in a *E. coli* whole-cell system with a range of different cyclic ketones (Anon 2007). Of all the monoxygenases discussed in this section, HAPMO seems to be the best candidate, due to the fact that there is already extensive literature about its substrate specificity and performance as an whole-cell biocatalyst. Additionally, it provides the lowest codon bias disparity with the *S. carnosus* host compared with the other Baeyer-Villiger monooxygenases, except for CHMO.

Another type of monooxygenase that is relevant for the study of whole-cell biocatalysis due to its catalytic importance and requirement for NAD(P)H and co-factor regeneration, is cytochrome P450s. Cytochrome P450s are heme containing oxygenases that have been found in most living organisms and are able to catalyse oxidation reactions a broad range of substrates, including steroids, prostaglandins, carcinogens and xenobiotics (Urlacher & Schmid 2002). Most P450s are mechanistically dependent on redox proteins such as a iron-sulfur ferrodoxin and flavin adenine dinucleotide (FAD)- or flavin mononucleotide (FMN)-reductases to perform the electron transfer from NAD(P)H (Guengerich 2007). This dependency on co-factor regeneration allied with better stability and activity rates when expressed intracellularly makes P450s very valuable candidates for whole-cell biocatalysis.

One of the most studied and characterized cytochrome P450s is the CYP102 from *Bacillus megaterium*, which is unique due to the fact that it is cytosolic, as opposed to membrane bound), and it is also self-sufficient, i.e. it contains both FNM/FAD reductase and monooxygenase domains in one single

chain (Narhi & Fulco 1986). Additionally, this enzyme has shown to have the highest reaction rates for reported activities in P450s, with turnover frequencies of more than 1000 min^{-1} (Schewe et al. 2008). These factors, together with the fact that the native host is a AT-rich Gram-positive bacteria, make the argument for CYP102 as an ideal candidate for expression in *S. carnosus*. Additionally hydrophobicity and related toxicity of long chain fatty acids and terpenoid substrates for CYP102, which are the most common limiting factors for biocatalysis with P450s (Urlacher & Eiben 2006), can be used as objective measurement in comparing the performances of *S. carnosus* and previously used recombinant *E. coli* strains (Schewe et al. 2008).

It is noteworthy to mention that the biocatalysts discussed in this section are a small sub-set from a much larger list of possible candidates. Our selection was done by choosing systems that seemed immediately relevant in the context of whole-cell systems, such as the dependency on co-factor regeneration, as well as the ease at which expression could be achieved in a recombinant bacterial strain. Thus, all the biocatalyst systems chosen were composed of a single gene of bacterial origin, and had been historically shown to be expressed in other recombinant bacterial systems. Consequently, there are many other candidate proteins that have been excluded from this selection that are probably more relevant for the biocatalytic industry. Regardless, we are confident that the selected set encloses a broad range of protein conformations, codon compositions, and substrate specificities that would allow for the study of several different aspects of whole-cell biocatalysis.

6.8.2- Towards the Creation of Universal Expression Systems

The construction of a vector system that could replicate and express proteins in Gram-positive and Gram-negative strains gave us the unique opportunity to test how promoters from the different strains could perform in foreign environments and interact with each other. Correspondingly, we found that the genomic promoters of *S. carnosus* are not exclusive systems that do not permit

protein expression in foreign systems. Thus, as a bio-product from the vector design strategy used in the project, we stumbled upon a new and alternative source for E. coli promoters in the genome of *S. carnosus*. We also found that genetic regulatory elements active in the staphylococcal genome could translate into a similar function in *E. coli.* These findings, which hint at a common genetic language between *E. coli* and *S. carnosus*, point toward the prospect of creating promoters that could operate in both bacterial strains, and even perhaps function in a broader range of Gram-negative and Gram-positive hosts as biocatalytic expression systems.

Correspondingly, one direction this project could take would be to elucidate the common transcription language between bacterial species by using the reporter vector pQR1031 as a genomic mining tool and to create 'universal' transcription systems that could work across Gram-negative and Gram-positive bacteria. In a similar approach to the optimization of promoter expression in *S. carnosus*, a library of different promoter constructs would be generated by mining both the staphylococcal and the *E. coli* genomes for regions upstream of house-keeping or well expressed genes. These genes are more likely to be conserved throughout the bacterial fauna due to their importance for survivability, and therefore are good targets as areas with a high level of genetic information conservation. Due to the interchangeable nature of the promoter system in pQR1031, the library of constructs would be easily produced by amplifying fragments from the genomes with *Eco*RI and *Xho*I sites at the ends, which would subsequently ligate with the vector cut with the same endonucleases. Additionally, the output of the different constructs could be easily assayable by measuring the tolerance level to gentamicin, thus providing an high-throughput method to screen a large promoter library.

The first stage of screening would be concerned with the selection of promoter areas that would provide a good base level for protein expression in both *E. coli* and *S. carnosus*, with the focus on constitutive expression. The results from this run would help us identify the elements within a promoter that are common between the bacterial species and can translate the same information regardless of expression hosts. The second stage would focus on the genetic analysis of elements in the proximity of the promoters that would carry the same functionality

across strains. This would involve the study of constructs that perhaps with low protein expression that could contain within their sequences elements that would potentially bind to TFs or form secondary structures, like cruciforms. Thus, this phase of the screening would allows us to build a catalogue of several regulatory elements that could be used interchangeably between bacterial species to produce similar inducible effects on protein expression systems.

While there have been several approaches focused on cataloging constitutive and inducible promoter systems for several bacterial species, such as Registry of Standard Biological Parts, there hasn't been to our knowledge any study conducted on the genomic level concerned with the cross-compatibility of promoters across different bacterial strains. Thus, we believe conducting such a study would add to the general understanding of the universal genetic language that can be used across different bacteria to promote and regulate gene transcription (Figure 78), while also adding to the current list of promoter modules that are used in the design of synthetic promoters.

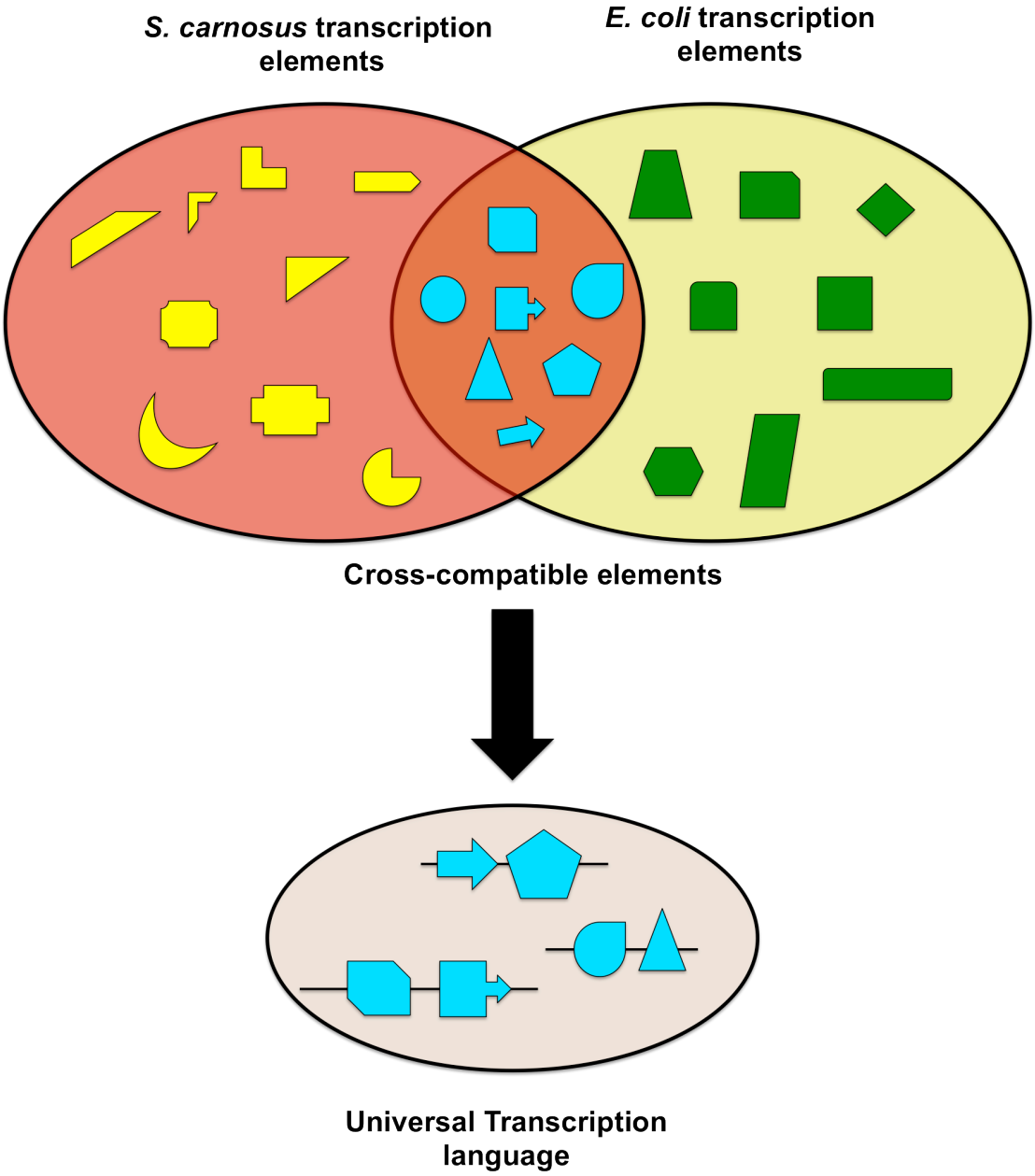


Figure 78- Principles underlying the construction of a library of universal promoter elements for use across bacterial species. Bacteria with different genetic biases use a majority of transcription elements that are unique to that bias. However, a small percentage of elements is cross-compatible between the species (blue shapes) , performing either the same or homologous functions in each of the strain. These intercepting points (blue shapes) can be used to construct a library of universal promoters that could be implemented in multiple bacterial strains.

6.8.3- Beyond the Biocatalytic Bubble: Adding to the Transcriptome of Staphylococcal Species

In the last sections, we described how the pQR1031 reporter vector could be used to develop novel promoter designs for biocatalytic expression within a bacterial strain and cross-species. However, the screening strategy proposed in these studies could be applied to fields outside biocatalysis. More specifically, the pQR1031 reporter vector could be applied to the area of bacterial pathology. The fact that *S. carnosus* is phylogenetically related to the pathogenic staphylococcal species, the most notorious and relevant to current health research being the methicillin-resistant *Staphylococcus aureus* (MRSA), makes the former a prime model system for the study of the biology underlying the pathology of these organisms and to test drug therapies against them.

Thus, by taking advantage of the non-pathogenicity of *S. carnosus* and the close homology between it's genome and that of pathogenic staphylococci, we could apply the screening strategies described above to map the activity and regulatory mechanisms of genomic promoters from the more pathogenic staphylococcal species.

There are already several powerful techniques that can be used to measure and map the levels of protein transcription (i.e the transcriptome) in bacterial genomes under different growth conditions. These include high-throughput methods such as DNA microarrays or Whole Transcriptome Shotgun Shotgun Sequencing (RNA-seq), both of which focus on the mRNA levels produced by cells. Transcriptomes are very valuable because they are a snapshot of genomic expression, and can be used to detect even slight changes in protein transcription under conditions that lead to pathogenicity of bacteria such as MRSA. However, while giving a general overview of the extracellular stimuli that affect mRNA levels in bacterial cells, transcriptomes do not communicate any information about the specific rates of transcription of the different proteins and the regulatory mechanisms that might be involved in this process and subsequent variations between transcription rates and translation rates.

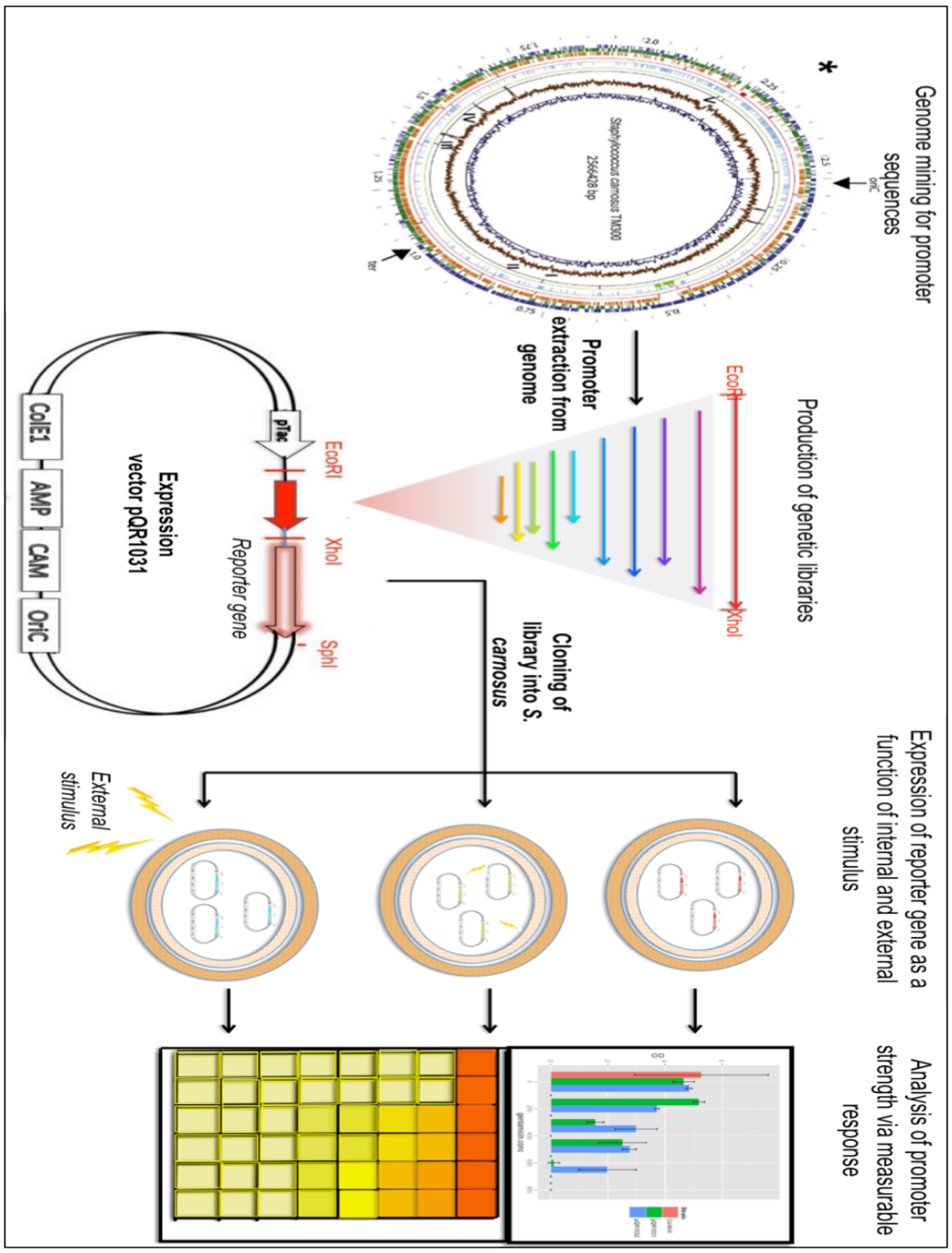


Figure 79- Genome mining strategy for the mapping of the promoter landscape of staphylococcal genomes. Sequences upstream of genes are lifted from several staphylococcal genomes and cloned into an hybrid vector with a interchangeable promoter region and a reporter gene. The Library of different promoter constructs is subsequently cloned into *S. carnosus* and promoter strength and regulation is analyzed by the phenotype of the reporter gene. In the case of the vector pQR1031, the level of expression would be measured as the level of tolerance to the antibiotic gentamicin (external stimulus)., as a function of growth in micro-well plates with increasing concentrations of the antibiotic. (*)- (Reproduced from the *International Journal of Medical Microbiology*, vol. 300 (2-3), Rosenstein & Gotz, Genomic differences between the food-grade S. carnosus and pathogenic staphylococcal species, pg. 104-108, Copyright © 2010, with permission from Elsevier.).

Thus we believe that by focusing on the genomic promoters, their activity and means of regulation in pathogenic staphylococcal strains would add to the current transcriptome data by mapping the levels of mRNA expression to specific genetic sequences and regulatory elements. This is not a novel approach to studying how genes are transcribed. Several groups have focused on specifically studying promoters of proteins that been involved in the pathogenicity of staphylococcal strains. However, as described previously, the interchangeability of the staphylococcal promoter in the pQR1031 reporter vector allows for an high-throughput approach in which a library of MSRA genomic sequences containing putative promoters could be easily constructed and characterized (Figure 79).

The characterization of a large population of genomic promoters would allow us to build a more in-depth picture of the different transcription systems involved in the pathogenicity of MRSA and perhaps highlight genetic targets for drug therapy. It is worth noting that the approach we are suggesting works under the assumption that *S. carnosus* and its pathogenic relatives employ similar transcription mechanisms, and that promoters lifted up from the pathogenic genomes would be under the influence and affected by the same factors in *S. carnosus*. This is perhaps a simplistic assumption, considering that some factors and proteins that trigger pathogenicity in *S. aureus* and *S. epidermis* are not expressed in *S. carnosus*. However, we believe that the majority of transcription factors and RNA polymerases involved in protein expression are equally present and preserved in most staphylococcus species, and as a result most promoters could be transcribed across different staphylococci.

7- References:

Alphand, V., Carrea, G., Wohlgemuth, R., Furstoss, R. & Woodley, J.M., 2003. Towards large-scale synthetic applications of Baeyer-Villiger monooxygenases. *Trends in biotechnology*, 21(7), pp.318–323.

Ayuso-Tejedor, S. Nishikori, S., Okuno, T. & Ogura, T., 2010. FtsH cleavage of non-native conformations of proteins. *Journal of structural biology*, 171(2), pp.117–124.

Baboo, J.Z., Galman, J.L., Lye, G.J., Ward, J.M., Hailes, H.C. & Micheletti, M., 2012. An automated microscale platform for evaluation and optimization of oxidative bioconversion processes. *Biotechnology Progress*, 28(2), pp.392–405.

Baldwin, C.V.F. & Woodley, J.M., 2006. On oxygen limitation in a whole cell biocatalytic Baeyer-Villiger oxidation process. *Biotechnology and Bioengineering*, 95(3), pp.362–369.

Becker, K., Bierbaum, G., von Eiff, C. & Engelmann, S., 2007. Understanding the physiology and adaptation of staphylococci: a post-genomic approach. *International journal of medical microbiology : IJMM*, 297(7-8), pp.483–501.

Beloqui, A., de Maria, P.D. & Golyshin, P.N., 2008. Recent trends in industrial microbiology. *Current opinion in microbiology*, 11(3), pp.240–248.

Bes, M.T., Villa, R., Roberts, S.M., Wan, P.W.H. & Willetts, A., 1996. Oxidative biotransformations by microorganisms: production of chiral synthons by cyclopentanone monooxygenase fromPseudomonas sp. NCIMB 9872. *Journal of Molecular Catalysis B: Enzymatic*, 1(3-6), pp.127–134.

Bikard, D., Loot, C., Baharoglu, Z. & Mazel, D., 2010. Folded DNA in Action: Hairpin Formation and Biological Functions in Prokaryotes. *Microbiology and Molecular Biology Reviews*, 74(4), pp.570–588.

Bommarius, A.S. & Riebel, B.R., 2005. Biocatalysis, Fundamentals and Applications. *Synthesis*, 2005(02), pp.338–338.

Bornscheuer, U.T. & Buchholz, K., 2005. Highlights in Biocatalysis – Historical Landmarks and Current Trends. *Engineering in Life Sciences*, 5(4), pp.309–323.

Breuer, M., Ditrich, K., Habicher, T., Hauer, B., KeBeler, M., Sturmer, R. & Zelinski, T., 2004. Industrial Methods for the Production of Optically Active Intermediates. *Angewandte Chemie International Edition*, 43(7), pp.788–824.

Burton, S.G., Cowan, D.A. & Woodley, J.M., 2002. The search for the ideal biocatalyst. *Nature biotechnology*, 20(1), pp.37–45.

Campbell, W.C., 2012. History of avermectin and ivermectin, with notes on the history of other macrocyclic lactone antiparasitic agents. *Current pharmaceutical biotechnology*, 13(6), pp.853–865.

Cheesman, M.J., Kneller, M.B., Kelly, E.J, Thompson, S.J., Yeung, C.K., Eaton, D.L. & Rettie, A.E.,

2001. Purification and Characterization of Hexahistidine-Tagged Cyclohexanone Monooxygenase Expressed in Saccharomyces cerevisiae and Escherichia coli. *Protein Expression and Purification*, 21(1), pp.81–86.

Chen, Y.C., Peoples, O.P. & Walsh, C.T., 1988. Acinetobacter cyclohexanone monooxygenase: gene cloning and sequence determination. *Journal of Bacteriology*, 170(2), pp.781–789.

Cho, B.K., Federowicz, S.A., Embree, M., Park, Y.S., Kim, D. & Palsson, B.O., 2011. The PurR regulon in Escherichia coli K-12 MG1655. *Nucleic Acids Research*, 39(15), pp.6456–6464.

Cohen, J.O., 1972. *The staphylococci*, John Wiley & Sons.

Coward-Kelly, G. & Chen, R.R., 2007. A window into biocatalysis and biotransformations. In Biotechnology progress. American Chemical Society, pp. 52–54.

Dayn, A., Malkhosyan, S., Duzhy, D., Lyamichev, V., Panchenko, Y. & Mirkin, S., 1991. Formation of (dA-dT)n cruciforms in Escherichia coli cells under different environmental conditions. *Journal of Bacteriology*, 173(8), pp.2658–2664.

Dayn, A., Malkhosyan, S. & Mirkin, S.M., 1992. Transcriptionally driven cruciform formation in vivo. *Nucleic Acids Research*, 20(22), pp.5991–5997.

de Bont, J., 1998. Solvent-tolerant bacteria in biocatalysis. *Trends in biotechnology*, 16(12), pp.493–499.

de Carvalho, C.C.C.R., 2011. Enzymatic and whole cell catalysis: finding new strategies for old processes. *Biotechnology advances*, 29(1), pp.75–83.

de Gonzalo, G., Mihovilovic, M.D. & Fraaije, M.W., 2010. Recent developments in the application of Baeyer-Villiger monooxygenases as biocatalysts. *ChemBioChem*, 11(16), pp.2208–2231.

de Vos, W.M., Kleerebezem, M. & Kuipers, O.P., 1997. Expression systems for industrial Gram-positive bacteria with low guanine and cytosine content. *Current opinion in biotechnology*, 8(5), pp.547–553.

Demain, A.L. & Sanchez, S., 2009. Microbial drug discovery: 80 years of progress. *The Journal of antibiotics*, 62(1), pp.5–16.

Demleitner, G. & Götz, F., 1994. Evidence for importance of the Staphylococcus hyicus lipase pro-peptide in lipase secretion, stability and activity. *FEMS Microbiology Letters*.

Dent, J.A., Davis, M.W. & Avery, L., 1997. avr-15 encodes a chloride channel subunit that mediates inhibitory glutamatergic neurotransmission and ivermectin sensitivity in Caenorhabditis elegans. *The EMBO Journal*, 16(19), pp.5867–5879.

Doig, S.D., Simpson, H., Alphand, V., Furstoss, R. & Woodley, J.M., 2003. Characterization of a recombinant Escherichia coli TOP10 [pQR239] whole-cell biocatalyst for stereoselective Baeyer-Villiger oxidations. *Enzyme and Microbial Technology*, 32(3-4), pp.347–355.

Doig, S.D., Avenell, P.J., Bird,P.A., Gallati, P., Lander, K.S., Lye, G.J., Wohlgemuth, R. & Woodley, J.M., 2002. Reactor operation and scale-up of whole cell Baeyer-Villiger catalyzed lactone synthesis. *Biotechnology Progress*, 18(5), pp.1039–1046.

Donald, L.J., Chernushevich, I.V., Zhou, J., Verentchikov, A., Poppe-Schriemer, N., Hosfield, D.J., Westmore, J.B., Ens, W., Duckworth, H.W. & Standing, K.G., 1996. Preparation and properties of pure, full-length IclR protein of Escherichia coli. Use of time-of-flight mass spectrometry to investigate the problems encountered. *Protein science : a publication of the Protein Society*, 5(8), pp.1613–1624.

Donald, L.J., Hosfield, D.J., Cuvelier, S.L., Ens, W., Standing, K.G. & Duckworth, H.W., 2001. Mass spectrometric study of the Escherichia coli repressor proteins, Ic1R and Gc1R, and their complexes with DNA. *Protein science : a publication of the Protein Society*, 10(7), pp.1370–1380.

Doo, E.H., Lee, W.H., Seo, H.S, Seo, J.H. & Park, J.B., 2009. Productivity of cyclohexanone oxidation of the recombinant Corynebacterium glutamicum expressing chnB of Acinetobacter calcoaceticus. *Journal of Biotechnology*, 142(2), pp.164–169.

Dordick, J.S., Marletta, M.A. & Klibanov, A.M., 1987. Polymerization of phenols catalyzed by peroxidase in nonaqueous media. *Biotechnology and Bioengineering*, 30(1), pp.31–36.

Escalante, A., Calderon, R., Valdivia, A., de Anda, R., Hernandez, G., Ramirez, O.T., Gosset, G. & Bolivar, F., 2010. Metabolic engineering for the production of shikimic acid in an evolved Escherichia coli strain lacking the phosphoenolpyruvate: carbohydrate phosphotransferase system. *Microbial cell factories*, 9(1), p.21.

Ferreira-Torres, C., Micheletti, M. & Lye, G.J., 2005. Microscale process evaluation of recombinant biocatalyst libraries: application to Baeyer–Villiger monooxygenase catalysed lactone synthesis. *Bioprocess and biosystems engineering*, 28(2), pp.83-93.

Fleming, A., 1979. On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of B. influenzae*. British journal of experimental pathology,* 60 (1), pp.3-13.

Fraaije, M.W. & Janssen, D.B,, 2007. Biocatalytic scope of Baeyer-Villiger monooxygenases. *Modern Biooxidation: Enzymes*.

Frees, D., Varmanen, P. & Ingmer, H., 2001. Inactivation of a gene that is highly conserved in Gram-positive bacteria stimulates degradation of non-native proteins and concomitantly increases stress tolerance in Lactococcus lactis. *Molecular microbiology*, 41(1), pp.93–103.

Frees, D., Qazi, S.N.A., Hill, P.J. & Ingmer, H., 2003. Alternative roles of ClpX and ClpP in Staphylococcus aureus stress tolerance and virulence. *Molecular microbiology*, 48(6), pp.1565–1578.

Frees, D., Savijoki, K., Varmanen, P. & Ingmer, H., 2007. Clp ATPases and ClpP proteolytic complexes regulate vital biological processes in low GC, Gram-positive bacteria. *Molecular microbiology*, 63(5), pp.1285–1295.

Gal, J., 2008. *The discovery of biological enantioselectivity: Louis Pasteur and the fermentation of tartaric acid, 1857--a review and analysis 150 yr later*, Wiley Subscription Services, Inc., A Wiley Company.

Gatermann, S. & Marre, R., 1989. Cloning and expression of Staphylococcus saprophyticus urease gene sequences in Staphylococcus carnosus and contribution of the enzyme to virulence.

Infection and Immunity, 57(10), pp.2998–3002.

Gehl, J., 2003. Electroporation: theory and methods, perspectives for drug delivery, gene therapy and research. *Acta physiologica Scandinavica*, 177(4), pp.437–447.

Glazer, A.N. & Nikaido, H., 2007. *Microbial Biotechnology*, Cambridge University Press.

Goldsmith, M. & Tawfik, D.S., 2012. Directed enzyme evolution: beyond the low-hanging fruit. *Current Opinion in Structural Biology*, 22(4), pp.406–412.

Gosset, G., Zhang, Z., Nayyar, S., Cuevas, W.A. & Saier, M.H., 2004. Transcriptome analysis of Crp-dependent catabolite control of gene expression in Escherichia coli. *Journal of Bacteriology*, 186(11), pp.3516–3524.

Götz, F.,Popp, F., Korn, E. & Schleifer, K.H., 1985. Complete nucleotide sequence of the lipase gene from Staphylococcus hyicus cloned in Staphylococcus carnosus. *Nucleic Acids Research*, 13(16), pp.5895–5906.

Götz, F., 2004. Staphylococci in colonization and disease: prospective targets for drugs and vaccines. *Current opinion in microbiology*, 7(5), pp.477–487.

Greiner, R. & Konietzny, U., 2006. EBSCOhost | 21534718 | Phytase for Food Application. *Food Technology & Biotechnology*.

Guengerich, F.P., 2007. Mechanisms of cytochrome P450 substrate oxidation: MiniReview. *Journal of Biochemical and Molecular Toxicology*, 21(4), pp.163–168.

Gupta, A. & Khare, S.K., 2009. Enzymes from solvent-tolerant microbes: useful biocatalysts for non-aqueous enzymology. *Critical reviews in biotechnology*, 29(1), pp.44–54.

Gustafsson, C., Govindarajan, S. & Minshull, J., 2004. Codon bias and heterologous protein expression. *Trends in biotechnology*, 22(7), pp.346–353.

Gvritishvili, A.G., Leung, K.W. & Tombran-Tink, J., 2010. Codon preference optimization increases heterologous PEDF expression. *PLOS ONE*, 5(11), p.e15056.

Gøtterup, J., Olsen, K., Knochel, S., Tjener, K., Stahnke, L.H. & Moller, J.K.S., 2008. Colour formation in fermented sausages by meat-associated staphylococci with different nitrite- and nitrate-reductase activities. *Meat Science*, 78(4), pp.492–501.

Haefner, S., Knietsch, A., Scholten, E., Braun, J., Lohscheidt, M. & Zelder, O., 2005. Biotechnological production and applications of phytases. *Applied microbiology and biotechnology*, 68(5), pp.588–597.

Hall, B.G., 2003. The EBG system of E. coli: origin and evolution of a novel ß-galactosidase for the metabolism of lactose. *Origin and Evolution of New Gene Functions*, 10(Chapter 5), pp.143–156.

Hall, B.G., Betts, P.W. & Wootton, J.C., 1989. DNA sequence analysis of artificially evolved ebg enzyme and ebg repressor genes. *Genetics*, 123(4), pp.635–648.

Hammes, W.P., 2012. Metabolism of nitrate in fermented meats: The characteristic feature of a

specific group of fermented foods. *Food Microbiology*, 29(2), pp.151–156.

Hansson, M., Samuelson, P., Nguyen, T.N., Stahl, S., 2002. General expression vectors for Staphylococcus carnosus enabled efficient production of the outer membrane protein A of Klebsiella pneumoniae. *FEMS Microbiology Letters*, 210(2), pp.263–270.

Hilker, I., Wohlgemuth, R., Alphand V. & Furstoss, R., 2005. Microbial transformations 59: First kilogram scale asymmetric microbial Baeyer-Villiger oxidation with optimized productivity using a resin-based in situ SFPR strategy. *Biotechnology and Bioengineering*, 92(6), pp.702–710.

Hilker, I., Gutierrez, M.C.G., Furstoss, R., Ward, J., Wohlgemuth, R. & Alphand, V., 2008. Preparative scale Baeyer|[ndash]|Villiger biooxidation at high concentration using recombinant Escherichia coli and in situ substrate feeding and product removal process. *Nature Protocols*, 3(3), pp.546–554.

Horinouchi, S. & Weisblum, B., 1982. Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *Journal of Bacteriology*, 150(2), pp.815–825.

Horwitz, M.S., 1989. Transcription regulation in vitro by an E. coli promoter containing a DNA cruciform in the "-35" region. *Nucleic Acids Research*, 17(14), pp.5537–5545.

Hueck, C.J., Hillen, W. & Saier, M.H., 1994. Analysis of a cis-active sequence mediating catabolite repression in gram-positive bacteria. *Research in Microbiology*, 145(7), pp.503–518.

Illanes, A., 2008. Enzyme Biocatalysis: Principles and Applications.

Illanes, A., Cauerhff, A., Wilson, L. & Castro G.R., 2012. Recent trends in biocatalysis engineering. *Bioresource Technology*, 115, pp.48–57.

Ishige, T., Honda, K. & Shimizu, S., 2005. Whole organism biocatalysis. *Current opinion in chemical biology*, 9(2), pp.174–180.

Iwaki, H., Hasegawa, Y., Wang, S., Kayser, M.M. & Lau, P.C.K., 2002. Cloning and characterization of a gene cluster involved in cyclopentanol metabolism in Comamonas sp. strain NCIMB 9872 and biotransformations effected by Escherichia coli-expressed cyclopentanone 1,2-monooxygenase. *Applied and Environmental Microbiology*, 68(11), pp.5671–5684.

JACOB, M., JAROS, D. & ROHM, H., 2011. Recent advances in milk clotting enzymes. *International Journal of Dairy Technology*, 64(1), pp.14–33.

Jankovic, I. & Brückner, R., 2002. Carbon catabolite repression by the catabolite control protein CcpA in Staphylococcus xylosus. *Journal of molecular microbiology and biotechnology*, 4(3), pp.309–314.

Jansen, R. & Bussemaker, R., 2003. Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. *Nucleic Acids Research*, 31(8), pp.2242–2251.

Jürgen, B., Hanschke, R., Sarvas, M., Hecker, M. & Schweder, T., 2001. Proteome and transcriptome based analysis of Bacillus subtilis cells overproducing an insoluble heterologous

protein. *Applied microbiology and biotechnology*, 55(3), pp.326–332.

Kamerbeek, N.M., Moonen, M.J., Van Der Ven, J.G., Van Berkel,W.J., Fraaije, M.W. & Janssen, D.B., 2001. 4-Hydroxyacetophenone monooxygenase from Pseudomonas fluorescens ACB. A novel flavoprotein catalyzing Baeyer-Villiger oxidation of aromatic compounds. *European journal of biochemistry / FEBS*, 268(9), pp.2547–2557.

Kamerbeek, N.M., Janssen, D.B., van Berkel, W.J.H., Fraaije, M.W., 2003. Baeyer–Villiger Monooxygenases, an Emerging Family of Flavin-Dependent Biocatalysts - Kamerbeek - 2003 - Advanced Synthesis & Catalysis - Wiley Online Library. *Advanced Synthesis & Catalysis*.

Kan, O., Griffiths, L., Baban, D., Iqball, S., Uden, M., Spearman, H., Slingsby, J., Price, T., Esapa, M., Kingsman, S., Kingsman, A., Slade, A. & Naylor, S., 2001. Direct retroviral delivery of human cytochrome P450 2B6 for gene-directed enzyme prodrug therapy of cancer. *Cancer gene therapy*, 8(7), pp.473–482.

Kardos, N. & Demain, A.L., 2011. Penicillin: the medicine with the greatest impact on therapeutic outcomes. *Applied microbiology and biotechnology*, 92(4), pp.677–687.

Kayser, M.M., 2009. “Designer reagents” recombinant microorganisms: new and powerful tools for organic synthesis. *Tetrahedron*, 65(5), pp.947–974.

Keller, G., Schleifer, K.H. & Götz, F., 1983. Construction and characterization of plasmid vectors for cloning in Staphylococcus aureus and Staphylococcus carnosus. *Plasmid*, 10(3), pp.270–278.

Khlebnikov, A., Skaug, T. & Keasling, J.D., 2002. Modulation of gene expression from the arabinose-inducible araBAD promoter. *Journal of Industrial Microbiology and Biotechnology*, 29(1), pp.34–37.

Kirk, O., Borchert, T.V. & Fuglsang, C.C., 2002. Industrial enzyme applications. *Current opinion in biotechnology*, 13(4), pp.345–351.

Kofman, A., Graf, M., Bojak, A., Deml, L., Bieler, K., Kharazova, A., Wolf, H. & Wagner, R., 2003. HIV-1 gag expression is quantitatively dependent on the ratio of native and optimized codons. *Tsitologiia*, 45(1), pp86-93

Kohler, C., Wolff, S., Albrecht, D., Fuchs, S., Becher, D., Buttner, K., Engelmann, S. & Hecker, M., 2005. Proteome analyses of Staphylococcus aureus in growing and non-growing cells: A physiological approach. *International Journal of Medical Microbiology*, 295(8), pp.547–565.

Kostichka, K., Thomas, S.M., Gibson, K.J., Nagarajan, V. & Cheng, Q., 2001. Cloning and Characterization of a Gene Cluster for Cyclododecanone Oxidation in Rhodococcus ruber SC1. *Journal of Bacteriology*, 183(21), pp.6478–6486.

Kraemer, G.R. & Iandolo, D.J.J., 1990. High-frequency transformation ofStaphylococcus aureus by electroporation. *Current Microbiology*, 21(6), pp.373–376.

Kuhn, D., Blank, L.M., Schmid, A. & Buhler, B., 2010. Systems biotechnology – Rational whole-cell biocatalyst and bioprocess design. *Engineering in Life Sciences*, 10(5), pp.384–397.

Lawson, C.L., Swigon, D., Murakami, K.S., Darst, S.A., Berman, H.M. & Ebright, R.H., 2004. Catabolite activator protein: DNA binding and transcription activation. *Current Opinion in*

Structural Biology, 14(1), pp.10–20.

Le Loir, Y., Azevedo, V., Oliveira, S.C., Freitas, D.A., Miyoshi, A., Bermudez-Humaran, L.G., Nouaille, S., Ribeiro, L.A., Leclercq, S., Gabriel, J.E., Guimaraes, V.D., Oliveira, M.N., Charlier, C., Gautier, M. & Langella, P., 2005. Protein secretion in Lactococcus lactis : an efficient way to increase the overall heterologous protein production. *Microbial cell factories*, 4(1), p.2.

Lee, S.Y., 2009. *Systems Biology and Biotechnology of Escherichia coli*, Springer.

Leon, R., Fernandes, P., Pinheiro, H.M. & Cabral., J.M.S.,1998. Whole-cell biocatalysis in organic media. *Enzyme and Microbial Technology*, 23(7-8), pp.483–500.

Li, W., Zhou, X. & Lu, P., 2004. Bottlenecks in the expression and secretion of heterologous proteins in Bacillus subtilis. *Research in Microbiology*, 155(8), pp.605–610.

Liese, A. & Villela Filho, M., 1999. Production of fine chemicals using biocatalysis. *Current opinion in biotechnology*, 10(6), pp.595–603.

Liese, A., Seelbach, K. & Wandrey, C., 2008. *Industrial Biotransformations*, John Wiley & Sons.

Liljeqvist, S., Samuelson, P., Hansson, M., Nguyen, T.N., Binz, H. & Stahl, S., 1997. Surface display of the cholera toxin B subunit on Staphylococcus xylosus and Staphylococcus carnosus. *Applied and Environmental Microbiology*, 63(7), pp.2481–2488.

Lin, L., Song, H., Ji, Y., He, Z., Pu, Y., Zhou, J. & Xu, J., 2010. Ultrasound-Mediated DNA Transformation in Thermophilic Gram-Positive Anaerobes. *PLOS ONE*, 5(9), p.e12582.

Lin, S.H. & Guidotti, G., 2009. Chapter 35 Purification of Membrane Proteins. In *Methods in Enzymology*. Methods in Enzymology. Elsevier, pp. 619–629.

Lorca, G.L., Chung, Y.J., Barabote, R.D., Weyler, W., Schilling, C.H. & Saier Jr, M.H., 2005. Catabolite Repression and Activation in Bacillus subtilis: Dependency on CcpA, HPr, and HprK. *Journal of Bacteriology*, 187(22), pp.7826–7839.

Lovett, P.S., 1996. Translation attenuation regulation of chloramphenicol resistance in bacteria — a review. *Gene*, 179(1), pp.157–162.

Löfblom, J., Kronqvist, N., Uhlen, M., Stahl, S. & Wernerus, H., 2007. Optimization of electroporation-mediated transformation: Staphylococcus carnosus as model organism. *Journal of Applied Microbiology*, 102(3), pp.736–747.

Lye, G., 1999. Application of in situ product-removal techniques to biocatalytic processes. *Trends in biotechnology*, 17(10), pp.395–402.

Lye, G.J., Shamlou, P.A., Baganz, F., Dalby, P.A. & Woodley, J.M., 2003. Accelerated design of bioconversion processes using automated microscale processing techniques. *Trends in biotechnology*, 21(1), pp.29–37.

Madsen, S.M., Beck, H.C., Ravn, P., Vrang, A., Hansen, A.M. & Israelsen, H., 2002. Cloning and Inactivation of a Branched-Chain-Amino-Acid Aminotransferase Gene from Staphylococcus carnosus and Characterization of the Enzyme. *Applied and Environmental Microbiology*, 68(8), pp.4007–4014.

Malito, E., Alfieri, A., Fraaije, M.W., Mattevi, A., 2004. Crystal structure of a Baeyer–Villiger monooxygenase. *Proceedings of the National Academy of Sciences of the United States of America*, 101(36), pp.13157–13162.

Masson, F., Hinrichsen, L., Talon, R. & Montel, M.C., 1999. Factors influencing leucine catabolism by a strain of Staphylococcus carnosus. *International Journal of Food Microbiology*, 49(3), pp.173–178.

MAURER, K., 2004. Detergent proteases. *Current opinion in biotechnology*, 15(4), pp.330–334.

Meijer, W.J.J. & Salas, M., 2004. Relevance of UP elements for three strong Bacillus subtilis phage φ29 promoters. *Nucleic Acids Research*, 32(3), pp.1166–1176.

Michel, A., Agerer, F., Hauck, C.R., Herrmann, M., Ullrich, J., Hacker, J. & Ohlsen, K., 2006. Global Regulatory Impact of ClpP Protease of Staphylococcus aureus on Regulons Involved in Virulence, Oxidative Stress Response, Autolysis, and DNA Repair. *Journal of Bacteriology*, 188(16), pp.5783–5796.

Mihovilovic, M.D., Rudroff, F., Kandioller, W., Grotzl, B., Stanetty, P. & Spreitzer, H., 2003. Synthesis and Enantioselective Baeyer-Villiger Oxidation of Prochiral Perhydro-pyranones with Recombinant E. coli Producing Cyclohexanone Monooxygenase. *Synlett*, 2003(13), pp.1973–1976.

Mihovilovic, M.D., Müller, B. & Stanetty, P., 2002. Monooxygenase-Mediated Baeyer−Villiger Oxidations. *European Journal of Organic Chemistry*, 2002(22), pp.3711–3730.

Mirza, I.A., Yachnin, B.J., Wang, S., Grosse, S., Bergeron, H., Imura, A., Iwaki, H., Hasegawa, Y., Lau, P.C.K. & Berghuis, A.M., 2009. Crystal structures of cyclohexanone monooxygenase reveal complex domain movements and a sliding cofactor. *Journal of the American Chemical Society*, 131(25), pp.8848–8854.

Monod, J., 2003. The Growth of Bacterial Cultures. *Annual Reviews in Microbiology*.

Montel, M.C., Masson, F. & Talon, R., 1998. Bacterial role in flavour development. *Meat Science*, 49, pp.S111–S123.

Morii, S., Sawamoto, S., Yamauchi, Y., Miyamoto, M., Iwami, M. & Itagaki, E., 1999. Steroid Monooxygenase of Rhodococcus rhodochrous: Sequencing of the Grenomic DNA, and Hyperexpression, Purification, and Characterization of the Recombinant Enzyme. *Journal of Biochemistry*, 126(3), pp.624–631.

Narhi, L.O. & Fulco, A.J., 1986. Characterization of a catalytically self-sufficient 119,000-dalton cytochrome P-450 monooxygenase induced by barbiturates in Bacillus megaterium. *The Journal of biological chemistry*, 261(16), pp.7160–7169.

Newman, D.J. & Cragg, G.M., 2012. Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *Journal of natural products*.

Ni, Y. & Chen, R.R., 2004. Accelerating whole-cell biocatalysis by reducing outer membrane permeability barrier. *Biotechnology and Bioengineering*, 87(6), pp.804–811.

O'Sullivan, D.J. & Klaenhammer, T.R., 1993. Rapid Mini-Prep Isolation of High-Quality Plasmid DNA

from Lactococcus and Lactobacillus spp. *Applied and Environmental Microbiology*, 59(8), pp.2730–2733.

Olsen, K.N., Budde, B.B., Siegumfeldt, H., Rechinger, K.B., Jakobsen, M. & Ingmer, H., 2002. Noninvasive Measurement of Bacterial Intracellular pH on a Single-Cell Level with Green Fluorescent Protein and Fluorescence Ratio Imaging Microscopy. *Applied and Environmental Microbiology*, 68(8), pp.4145–4147.

Otten, L.G., Hollmann, F. & Arends, I.W.C.E., 2010. Enzyme engineering for enantioselectivity: from trial-and-error to rational design? *Trends in biotechnology*, 28(1), pp.46–54.

O'Sullivan, L.M., Patel, S., Ward, J.M., Woodley, J.M. & Doig, S.D., 2001. Large scale production of cyclohexanone monooxygenase from Escherichia coli TOP10 pQR239. *Enzyme and Microbial Technology*, 28(2-3), pp.265–274.

Panke, S. & Wubbolts, M., 2005. Advances in biocatalytic synthesis of pharmaceutical intermediates. *Current opinion in chemical biology*, 9(2), pp.188–194.

Pantůcek, R., Sedlacek, I., Doskar, J. & Rosypal, S., 1999. Complex genomic and phenotypic characterization of the related species Staphylococcus carnosus and Staphylococcus piscifermentans. *International Journal of Systematic Bacteriology*, 49 Pt 3, pp.941–951.

Park, J.-W., Lee, J.K., Kwon, T.J., Yi, D.H., Kim, Y.J., Moon, S.H., Suh, H.H., Kang, S.M. & Park, Y.I., 2003. Bioconversion of compactin into pravastatin by Streptomyces sp. *Biotechnology Letters*, 25(21), pp.1827–1831.

Patel, R., Hanson, R., Goswami, A., Nanduri, V., Banerjee, A., Donovan, M.J., Goldberg, S., Johnston, R., Brzozowski , D., Tully, T., Howell, J., Cazzulino, D. & Ko, R., 2003. Enzymatic synthesis of chiral intermediates for pharmaceuticals. *Journal of Industrial Microbiology & Biotechnology*, 30(5), pp.252-259.

Plotkin, J.B. & Kudla, G., 2010. Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics*, 12(1), pp.32–42.

Pollard, D.J. & Woodley, J.M., 2007. Biocatalysis for pharmaceutical intermediates: the future is now. *Trends in biotechnology*, 25(2), pp.66–73.

Pope, B. & Kent, H.M., 1996. High efficiency 5 min transformation of Escherichia coli. *Nucleic Acids Research*, 24(3), pp.536–537.

Puigbò, P., Bravo, I.G. & Garcia-Vallvé, S., 2008. CAIcal: A combined set of tools to assess codon usage adaptation. *Biology Direct*, 3(1), p.38.

Rahmouni, A.R. & Wells, R.D., 1992. Direct evidence for the effect of transcription on local DNA supercoiling in vivo. *Journal of Molecular Biology*, 223(1), pp.131–144.

Reggelin, M., 2005. Chemical Structure, Spatial Arrangement. The Early History of Stereochemistry, 1874–1914. Von Peter J. Ramberg. *Angewandte Chemie*, 117(5), pp.675–677.

Rehdorf, J., Mihovilovic, M.D., Fraaije, M.W. & Bornscheuer, U.T., 2010. Enzymatic Synthesis of Enantiomerically Pure β-Amino Ketones, β-Amino Esters, and β-Amino Alcohols with Baeyer–Villiger Monooxygenases. *Chemistry - A European Journal*, 16(31), pp.9525–9535.

Rhodius, V.A., Mutalik, V.K. & Gross, C.A., 2012. Predicting the strength of UP-elements and full-length E. coli σE promoters. *Nucleic Acids Research*, 40(7), pp.2907–2924.

Rosenstein, R., Nerz, C., Biswas, L., Resch, A., Raddatz, G., Schuster, SC. & Götz, F., 2009. Genome Analysis of the Meat Starter Culture Bacterium Staphylococcus carnosus TM300. *Applied and Environmental Microbiology*, 75(3), pp.811–822.

Rosenstein, R. & Götz, F., 2010. Genomic differences between the food-grade Staphylococcus carnosus and pathogenic staphylococcal species. *International Journal of Medical Microbiology*, 300(2-3), pp.104–108.

Samuelson, P., Hansson, M., Ahlborg, N.,Andreoni, C., Gotz, F., Bachi, T., Ngueyn, T.N., Binz, H., Uhlen, M. & Stahl, S., 1995. Cell surface display of recombinant proteins on Staphylococcus carnosus. *Journal of Bacteriology*, 177(6), pp.1470–1476.

Samuelson, P., Cano, F., Robert, A. & Stahl, S., 1999. Engineering of a Staphylococcus carnosus surface display system by substitution or deletion of a Staphylococcus hyicus lipase propeptide. *FEMS Microbiology Letters*, 179(1), pp.131–139.

Sanchez, S. & Demain, A.L., 2010. Enzymes and Bioconversions of Industrial, Pharmaceutical, and Biotechnological Significance. *Organic Process Research & …*.

Sardessai, Y. & Bhosle, S., 2002. Tolerance of bacteria to organic solvents. *Research in Microbiology*, 153(5), pp.263–268.

Schewe, H., Kaup, B.-A. & Schrader, J., 2008. Improvement of P450BM-3 whole-cell biocatalysis by integrating heterologous cofactor regeneration combining glucose facilitator and dehydrogenase in E. coli. *Applied microbiology and biotechnology*, 78(1), pp.55–65.

Schleifer, K.H. & Fischer, U., 1982. Description of a New Species of the Genus Staphylococcus: Staphylococcus carnosus. *International Journal of Systematic Bacteriology*, 32(2), pp.153–156.

Schmid, A., Dordick, J.S., Hauer, B., Hauer, B., Kiener, A., Wubbolts, M. & Witholt, B., 2001. Industrial biocatalysis today and tomorrow. *Nature*, 409(6817), pp.258–268.

Schoemaker, H., Mink, D. & Wubbolts, M., 2003. Dispelling the myths--biocatalysis in industrial synthesis. *Science*, 299(5613), pp.1694–1697.

Schofield, D.A., Westwater, C., Hoel, B.D., Werner, P.A., Norris, S.J. and Schmidt, M.G., 2003. Development of a Thermally Regulated Broad-Spectrum Promoter System for Use in Pathogenic Gram-Positive Species. *Applied and Environmental Microbiology*, 69(6), pp.3385–3392.

Schroer, K., Kittelmann, M. & Lütz, S., 2010. Recombinant human cytochrome P450 monooxygenases for drug metabolite synthesis. *Biotechnology and Bioengineering*, 106(5), pp.699–706.

Seidl, K., Muller, S., Francois, P., Kriebitzsch, C., Schrenzel, J., Engelmann, S., Bischoff, M. & Berger-Bachi, B., 2009. Effect of a glucose impulse on the CcpA regulon in Staphylococcus aureus. *BMC microbiology*, 9(1), p.95.

Seymour, A.A., Swerdel, J.N. & Abboa-Offei, B., 1991. Antihypertensive Activity During Inhibition of

Neutral Endopeptidase and Angiotensin Converting Enzyme. *Journal of Cardiovascular Pharmacology*, 17(3), p.456.

Sharp, P.M. & Li, W.H., 1987. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*, 15(3), pp.1281–1295.

Sharp, P.M., Bailes, E., Grocock, R.J., Peden, J.F. & Sockett, R.E., 2005. Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Research*, 33(4), pp.1141–1153.

Shaw, A.J., IV, 2008. *Metabolic Engineering of High Yield Ethanol Production in Thermoanaerobacterium Saccharolyticum*, ProQuest.

Shieh, C.-J., Phan Thi, L.-A. & Shih, I.-L., 2009. Milk-clotting enzymes produced by culture of Bacillus subtilis natto. *Biochemical Engineering Journal*, 43(1), pp.85–91.

Sikkema, J., de Bont, J.A. & Poolman, B., 1995. Mechanisms of membrane toxicity of hydrocarbons. *Microbiology and Molecular Biology Reviews*, 59(2), pp.201–222.

Simonová, M., Strompfova, V., Marcinakova, M., Laukova, A., Vesterlund, S., Moratalla, M.L., Bover-Cid, S. & Vidal-Carou, C., 2006. Characterization of Staphylococcus xylosus and Staphylococcus carnosus isolated from Slovak meat products. *Meat Science*, 73(4), pp.559–564.

Simpson, H.D., Alphand, V. & Furstoss, R., 2001. Microbiological transformations. *Journal of Molecular Catalysis B: Enzymatic*, 16(2), pp.101–108.

Singh, J., Mukerji, M. & Mahadevan, S., 1995. Transcriptional activation of the Escherichia coli bgl operon: negative regulation by DNA structural elements near the promoter. *Molecular microbiology*, 17(6), pp.1085–1092.

Singh, R.K., Tiwari, M.K., Singh, R. & Lee, J.K., 2013. From Protein Engineering to Immobilization: Promising Strategies for the Upgrade of Industrial Enzymes. *International journal of molecular …*.

Stark, M.J.R., 1987. Multicopy expression vectors carrying the lac represser gene for regulated high-level expression of genes in Escherichia coli. *Gene*, 51(2-3), pp.255–267.

Stewart, J.D., Reed, K.W. & Kayser, M.M., 1996. “Designer yeast”: a new reagent for enantioselective Baeyer–Villiger oxidations. *J. Chem. Soc., Perkin Trans. 1*, (8), pp.755–757.

Suzuki, H., Lefebure, T., Bitar, P.P. & Stanhope, M.J., 2012. Comparative genomic analysis of the genus Staphylococcus including Staphylococcus aureus and its newly described sister species Staphylococcus simiae. *BMC Genomics*, 13(1), p.38.

Swint-Kruse, L. & Matthews, K.S., 2009. Allostery in the LacI/GalR family: variations on a theme. *Current opinion in microbiology*, 12(2), pp.129–137.

Søndergaard, A.K. & Stahnke, L.H., 2002. Growth and aroma production by Staphylococcus xylosus, S. carnosus and S. equorum—a comparative study in model systems. *International Journal of Food Microbiology*, 75(1-2), pp.99–109.

Talon, R. & Montel, M.C., 1997. Hydrolysis of esters by staphylococci. *International Journal of Food*

Microbiology, 36(2-3), pp.207–214.

Talon, R., Chastagnac, C., Vergnais, L., Montel, M.C. & Berdague, J.L., 1998. Production of esters by Staphylococci. *International Journal of Food Microbiology*, 45(2), pp.143–150.

Talon, R., Walter, D., Chartier, S., Barriere, C. & Montel, M.C., 1999. Effect of nitrate and incubation conditions on the production of catalase and nitrate reductase by staphylococci. *International Journal of Food Microbiology*, 52(1-2), pp.47–56.

Thomas, S.M., DiCosimo, R. & Nagarajan, V., 2002. Biocatalysis: applications and potentials for the chemical industry. *Trends in biotechnology*, 20(6), pp.238–242.

Titgemeyer, F. & Hillen, W., 2002. Global control of sugar metabolism: a gram-positive solution. *Antonie van Leeuwenhoek*, 82(1-4), pp.59–71.

Tjener, K., Stahnke, L.H., Andersen, L. & Martinussen, J., 2004. Growth and production of volatiles by Staphylococcus carnosus in dry sausages: Influence of inoculation level and ripening time. *Meat Science*, 67(3), pp.447–452.

Torres Pazmiño, D.E. et al., 2009. Efficient Biooxidations Catalyzed by a New Generation of Self-Sufficient Baeyer–Villiger Monooxygenases. *ChemBioChem*, 10(16), pp.2595–2598.

Tsuchiya, Y., Nakajima, M. & Yokoi, T., 2005. Cytochrome P450-mediated metabolism of estrogens and its regulation in human. *Cancer letters*, 227(2), pp.115–124.

Tzanov, T., Calafell, M., Guebitz, G.M. & Cavaco-Paulo, A., 2001. Bio-preparation of cotton fabrics. *Enzyme and Microbial Technology*, 29(6-7), pp.357–362.

Urlacher, V. & Schmid, R.D., 2002. Biotransformations using prokaryotic P450 monooxygenases. *Current opinion in biotechnology*, 13(6), pp.557–564.

Urlacher, V.B. & Eiben, S., 2006. Cytochrome P450 monooxygenases: perspectives for synthetic application. *Trends in biotechnology*, 24(7), pp.324–330.

Urlacher, V.B. & Girhard, M., 2012. Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. *Trends in biotechnology*, 30(1), pp.26–36.

van Beilen, J.B., Duetz, W.A., Schmid, A. & Witholt, B., 2003. Practical issues in the application of oxygenases. *Trends in biotechnology*, 21(4), pp.170–177.

Vishwanatha, K.S., Appu Rao, A.G. & Singh, S.A., 2010. Production and characterization of a milk-clotting enzyme from Aspergillus oryzae MTCC 5341. *Applied microbiology and biotechnology*, 85(6), pp.1849–1859.

Vollmer, W., Blanot, D. & de Pedro, M.A., 2008. Peptidoglycan structure and architecture. *FEMS Microbiology Reviews*, 26(15), pp.3584-3590.

Vollmer, W. & Seligman, S.J., 2010. Architecture of peptidoglycan: more data and more models. *Trends in Microbiology*, 18(2), pp.59–66.

Voskuil, M., 1998. The -16 region of Bacillus subtilis and other gram-positive bacterial promoters. *Nucleic Acids Research*, 26(15), pp.3584–3590.

Waites, M.J., Morgan, N.L., Rockey, J.S. & Higton, G., 2009. *Industrial Microbiology*, John Wiley & Sons.

Walton, A.Z. & Stewart, J.D., 2002. An Efficient Enzymatic Baeyer–Villiger Oxidation by Engineered Escherichiacoli Cells under Non-Growing Conditions. *Biotechnology Progress*, 18(2), pp.262–268.

Walton, A.Z. & Stewart, J.D., 2004. Understanding and Improving NADPH-Dependent Reactions by Nongrowing Escherichia coli Cells. *Biotechnology Progress*, 20(2), pp.403–411.

Wang, J.C. & Lynch, A.S., 1993. Transcription and DNA supercoiling. *Current Opinion in Genetics & Development*, 3(5), pp.764–768.

Webb, J.S., Barratt, S.R., Sabev, H., Nixon, M., Eastwood, I.M., Greenhalgh, M., Handly, P.S. & Robson, G.D., 2001. Green Fluorescent Protein as a Novel Indicator of Antimicrobial Susceptibility in Aureobasidium pullulans. *Applied and Environmental Microbiology*, 67(12), pp.5614–5620.

Wieland, K.-P., Wieland, B. & Götz, F., 1995. A promoter-screening plasmid and xylose-inducible, glucose-repressible expression vectors for Staphylococcus carnosus. *Gene*, 158(1), pp.91–96.

Willetts, A., 1997. Structural studies and synthetic applications of Baeyer-Villiger monooxygenases. *Trends in biotechnology*, 15(2), pp.55–62.

Williams, R.J., Nair, S.P., Henderson, B., Holland, K.T., & Ward, J.M., 2002. Expression of the S. aureus hysA gene in S. carnosus from a modified E. coli–staphylococcal shuttle vector. *Plasmid*, 47(3), pp.241–245.

Wohlgemuth, R., 2010. Asymmetric biocatalysis with microbial enzymes and cells. *Current opinion in microbiology*, 13(3), pp.283–292.

Woodley, J.M., Bisschops, M., Straathof, A.J.J. & Ottens, M., 2008. Future directions for in-situ product removal (ISPR). *Journal of Chemical Technology and Biotechnology*, 83(2), pp.121–123.

Woodruff, M.A. & Hutmacher, D.W., 2010. The return of a forgotten polymer—Polycaprolactone in the 21st century. *Progress in Polymer Science*, 35(10), pp.1217–1256.

Wu, J.A., Kusuma, C., Mond, J.J. & Kokai-kun, J.F., 2003. Lysostaphin Disrupts Staphylococcus aureus and Staphylococcus epidermidis Biofilms on Artificial Surfaces. *Antimicrobial Agents and Chemotherapy*, 47(11), pp.3407–3414.

Xia, G., Kohler, T. & Peschel, A., 2010. The wall teichoic acid and lipoteichoic acid polymers of Staphylococcus aureus. *International Journal of Medical Microbiology*, 300(2-3), pp.148–154.

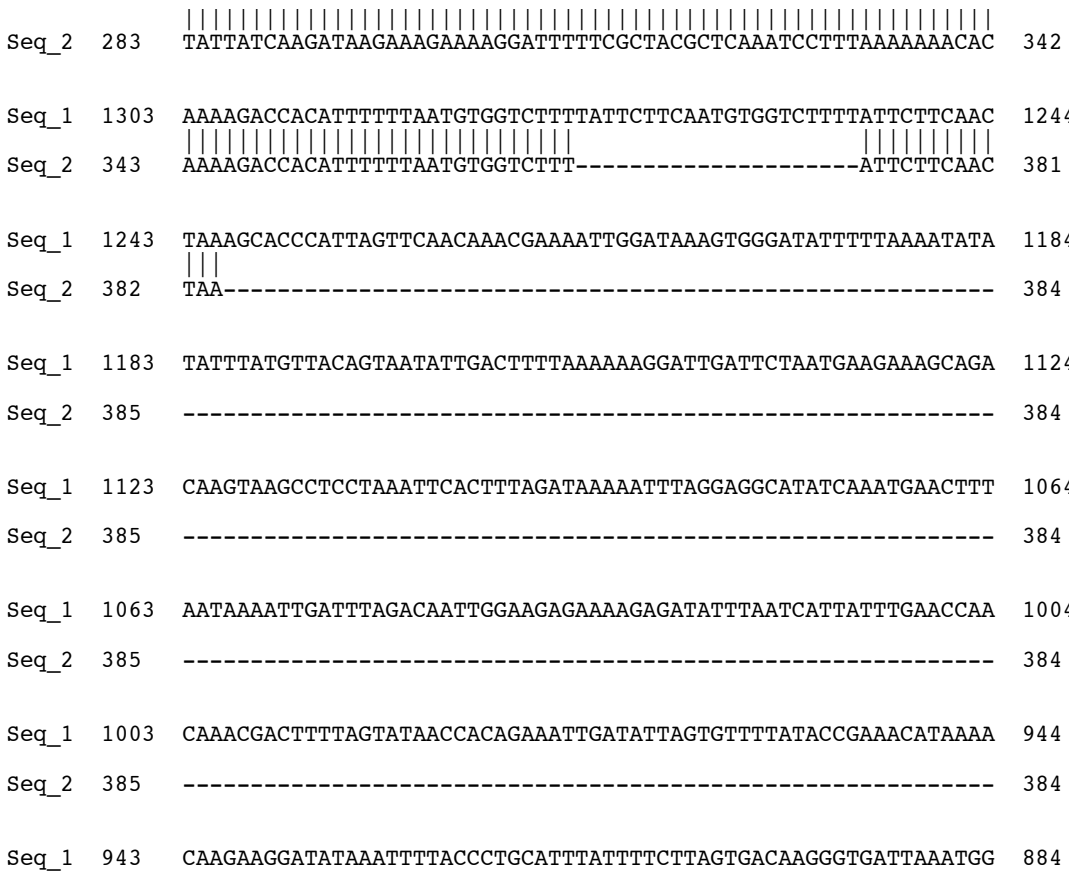
Yang, J., Wang, S. & Lorrain, M.J., 2009. Bioproduction of lauryl lactone and 4-vinyl guaiacol as value-added chemicals in two-phase biotransformation systems. *Applied Microbiology and Biotechnology*, 84(5), pp.867-876.

Yang, S.T., 2011. *Bioprocessing for Value-Added Products from Renewable Resources*, Elsevier.

Appendix

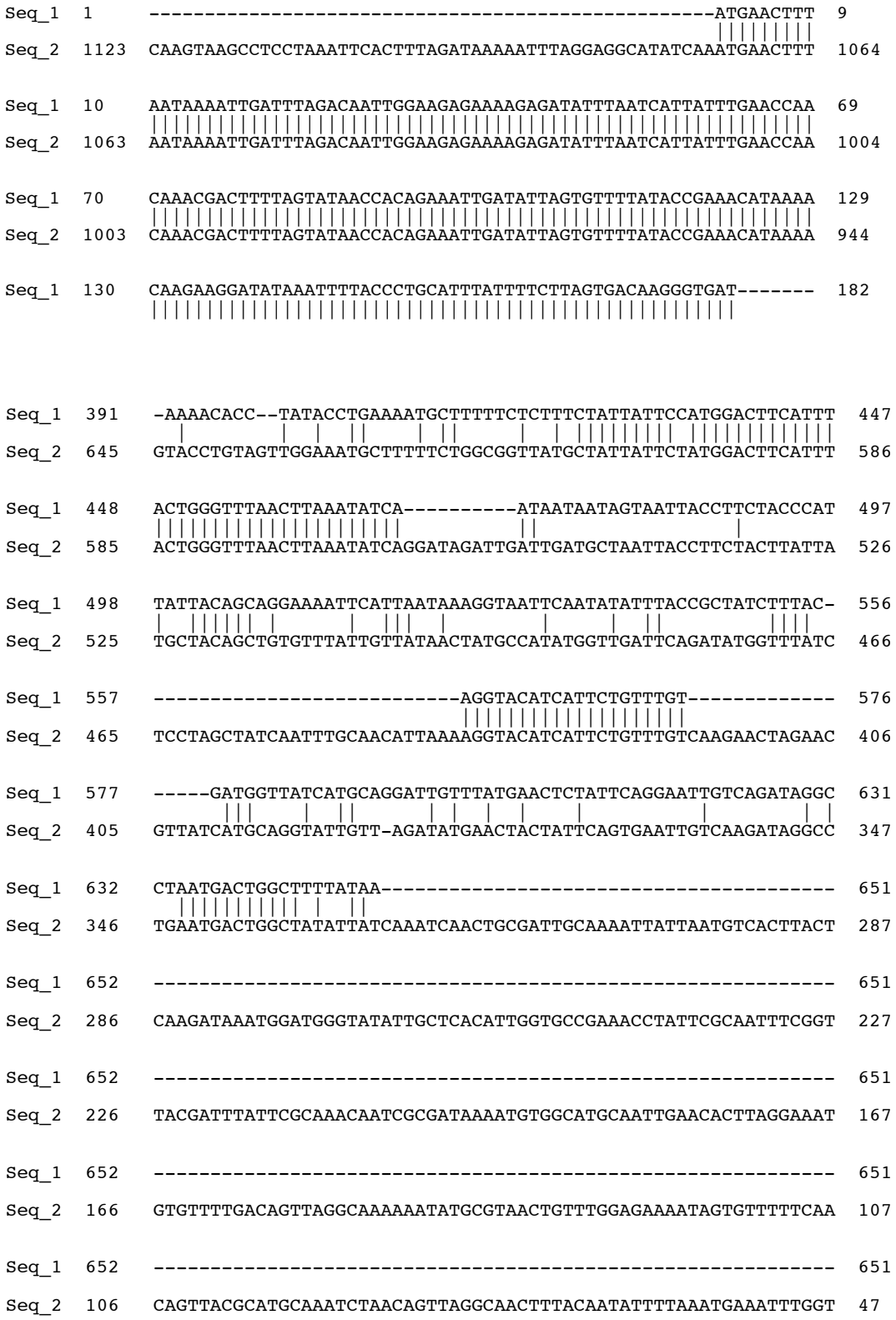
Appendix 1: Sequencing results

- Comparison between oriS of pC194 and sequencing from pNW21



Where Seq_1 is the sequenced pNW21 fragment, and Seq_2 is the oriS from pC194.

- Comparison between the CAM of pC194 and sequencing from pNW21



Where Seq_1 is the CAM sequence pC194, and Seq_2 is the sequenced pNW21 fragment.

- Comparison between sequencing results and the predicted sequence for pQR1030

```
Seq_1   28    ACACTAAAAAAGCAAAATAGTATTGATTTTACATTTTTTAAATGATATAATACTGTGGTC  87
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   361   ACACTAAAAAAGCAAAATAGTATTGATTTTACATTTTTTAAATGATATAATACTGTGGTC  420

Seq_1   88    GTGCTCGTAAAGGGTAGGCCATTTCGTCACGAAATGTTTTATGAGTGGGAGGGCAAAAAT  147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   421   GTGCTCGTAAAGGGTAGGCCATTTCGTCACGAAATGTTTTATGAGTGGGAGGGCAAAAAT  480

Seq_1   148   GAGCCCTGTGACCACATCACGATATCAACTCGGGGGTACCCTGAGATGTCACAAAAAATG  207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   481   GAGCCCTGTGACCACATCACGATATCAACTCG-----------AGATGTCACAAAAAATG  529

Seq_1   208   GATTTTGATGCTATCGTGATTGGTGGTGGTTTTGGCGGACTTTATGCAGTCAAAAAATTA  267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   530   GATTTTGATGCTATCGTGATTGGTGGTGGTTTTGGCGGACTTTATGCAGTCAAAAAATTA  589

Seq_1   268   AGAGACGAGCTCGAACTTAAGGTTCAGGCTTTTGATAAAGCCACGGATGTCGCAGGTACT  327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   590   AGAGACGAGCTCGAACTTAAGGTTCAGGCTTTTGATAAAGCCACGGATGTCGCAGGTACT  649

Seq_1   328   TGGTACTGGAACCGTTACCCAGGTGCATTGACGGATACAGAAACCCACCTCTACTGCTAT  387
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   650   TGGTACTGGAACCGTTACCCAGGTGCATTGACGGATACAGAAACCCACCTCTACTGCTAT  709

Seq_1   388   TCTTGGGATAAAGAATTACTACAATCGCTAGAAATCAAGAAAAAATATGTGCAAGGCCCT  447
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2   710   TCTTGGGATAAAGAATTACTACAATCGCTAGAAATCAAGAAAAAATATGTGCAAGGCCCT  769

Seq_1   448   GATGTACGCAAGTATTTACAGCAAGTGGCTGAAAAGCATGATTTAAAGAAGAGCTATCAA  507
```

Where Seq_1 is the sequenced pQR1030 fragment, and Seq_2 is the predicted sequence for pQR1030 around the CHMO gene.

- Comparison between sequencing results and the predicted sequence for pQR1032

```
Seq_1  1     ----------------------------------GAATTCAAAATAGTATTGATTTTACA  26
                                               ||||||||||||||||||||||||||
Seq_2  301   GAGCGGATAACAATTTCACACAGGAAACAGCGATGAATTCAAAATAGTATTGATTTTACA  360

Seq_1  27    TTTTTTAAATGATATAATACTGTGGTCGTGCTCGTAGGAGGTACTCGAGATGTTACGCAG  86
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  361   TTTTTTAAATGATATAATACTGTGGTCGTGCTCGTAGGAGGTACTCGAGATGTTACGCAG  420

Seq_1  87    CAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAACAAAGTTAGGTGGCTCAAGTAT  146
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  421   CAGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAACAAAGTTAGGTGGCTCAAGTAT  480

Seq_1  147   GGGCATCATTCGCACATGTAGGCTCGGCCCTGACCAAGTCAAATCCATGCGGGCTGCTCT  206
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  481   GGGCATCATTCGCACATGTAGGCTCGGCCCTGACCAAGTCAAATCCATGCGGGCTGCTCT  540

Seq_1  207   TGATCTTTTCGGTCGTGAGTTCGGAGACGTAGCCACCTACTCCCAACATCAGCCGGACTC  266
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  541   TGATCTTTTCGGTCGTGAGTTCGGAGACGTAGCCACCTACTCCCAACATCAGCCGGACTC  600

Seq_1  267   CGATTACCTCGGGAACTTGCTCCGTAGTAAGACATTCATCGCGCTTGCTGCCTTCGACCA  326
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  601   CGATTACCTCGGGAACTTGCTCCGTAGTAAGACATTCATCGCGCTTGCTGCCTTCGACCA  660

Seq_1  327   AGAAGCGGTTGTTGGCGCTCTCGCGGCTTACGTTCTGCCCAGGTTTGAGCAGCCGCGTAG  386
             |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Seq_2  661   AGAAGCGGTTGTTGGCGCTCTCGCGGCTTACGTTCTGCCCAAGTTTGAGCAGCCGCGTAG  720

Seq_1  387   TGAGATCTATATCTATGATCTCGCAGTCTCCGGCGAGCACCGGAGGCAGGGCATTGCCAC  446
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  721   TGAGATCTATATCTATGATCTCGCAGTCTCCGGCGAGCACCGGAGGCAGGGCATTGCCAC  780

Seq_1  447   CGCGCTCATCAATCTCCTCAAGCATGAGGCCAACGCGCTTGGTGCTTATGTGATCTACGT  506
```

Where Seq_1 is the predicted pQR1032 sequence around the rplK promoter and GenR, and Seq_2 is the sequencing result

- Comparison between sequencing results and the predicted sequence for pQR1033

```
Seq_1  30    TCACATCCAAACCGTTATATTGGTTAATGTGATTAGGAGGTTCTCGAGATGTTACGCAGC  89
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Seq_2  361   TCACATCCAAACCGTTATATTGGTTAAT-TGATTAGGAGGTTCTCGAGATGTTACGCAGC  419

Seq_1  90    AGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAACAAAGTTAGGTGGCTCAAGTATG  149
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  420   AGCAACGATGTTACGCAGCAGGGCAGTCGCCCTAAAACAAAGTTAGGTGGCTCAAGTATG  479

Seq_1  150   GGCATCATTCGCACATGTAGGCTCGGCCCTGACCAAGTCAAATCCATGCGGGCTGCTCTT  209
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  480   GGCATCATTCGCACATGTAGGCTCGGCCCTGACCAAGTCAAATCCATGCGGGCTGCTCTT  539

Seq_1  210   GATCTTTTCGGTCGTGAGTTCGGAGACGTAGCCACCTACTCCCAACATCAGCCGGACTCC  269
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  540   GATCTTTTCGGTCGTGAGTTCGGAGACGTAGCCACCTACTCCCAACATCAGCCGGACTCC  599

Seq_1  270   GATTACCTCGGGAACTTGCTCCGTAGTAAGACATTCATCGCGCTTGCTGCCTTCGACCAA  329
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  600   GATTACCTCGGGAACTTGCTCCGTAGTAAGACATTCATCGCGCTTGCTGCCTTCGACCAA  659

Seq_1  330   GAAGCGGTTGTTGGCGCTCTCGCGGCTTACGTTCTGCCCAGGTTTGAGCAGCCGCGTAGT  389
             |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Seq_2  660   GAAGCGGTTGTTGGCGCTCTCGCGGCTTACGTTCTGCCCAAGTTTGAGCAGCCGCGTAGT  719

Seq_1  390   GAGATCTATATCTATGATCTCGCAGTCTCCGGCGAGCACCGGAGGCAGGGCATTGCCACC  449
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  720   GAGATCTATATCTATGATCTCGCAGTCTCCGGCGAGCACCGGAGGCAGGGCATTGCCACC  779

Seq_1  450   GCGCTCATCAATCTCCTCAAGCATGAGGCCAACGCGCTTGGTGCTTATGTGATCTACGTG  509
```

Where Seq_1 is the predicted pQR1033 sequence around the rplJ promoter and GenR, and Seq_2 is the sequencing result.

- Comparison between sequencing results and the predicted sequence for pQR1034

```
Seq_1  103   ATCGTGATTGGTGGTGGTTTTGGCGGACTTTATGCAGTCAAAAAATTAAGAGACGAGCTC  162
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  541   ATCGTGATTGGTGGTGGTTTTGGCGGACTTTATGCAGTCAAAAAATTAAGAGACGAGCTC  600

Seq_1  163   GAACTTAAGGTTCAGGCTTTTGATAAAGCCACGGATGTCGCAGGTACTTGGTACTGGAAC  222
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  601   GAACTTAAGGTTCAGGCTTTTGATAAAGCCACGGATGTCGCAGGTACTTGGTACTGGAAC  660

Seq_1  223   CGTTACCCAGGTGCATTGACGGATACAGAAACCCACCTCTACTGCTATTCTTGGGATAAA  282
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  661   CGTTACCCAGGTGCATTGACGGATACAGAAACCCACCTCTACTGCTATTCTTGGGATAAA  720

Seq_1  283   GAATTACTACAATCGCTAGAAATCAAGAAAAAATATGTGCAAGGCCCTGATGTACGCAAG  342
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Seq_2  721   GAATTACTACAATCGCTAGAAATCAAGAAAAAATATGTGCAAGGCCCTGATGTACGCAAG  780

Seq_1  343   TATTTACAGCAAGTGGCTGAAAAGCATGATTTAAAGAAGAGCTATCAATTCAATACCGCG  402
```

```
Seq_1  1003  AAAGATCCAGCCATTGCACAGAAGCTTATGCCACAGGATTTGTATGCAAAACGTCCGTTG  1062
Seq_2  1061  ------------------------------------------------------------  1060

Seq_1  1063  TGTGACAGTGGTTACTACAACACCTTTAACCGTGACAATGTCCGTTTAGAAGATGTGAAA  1122
Seq_2  1061  ------------------------------------------------------------  1060

Seq_1  1123  GCCAATCCGATTGTTGAAATTACCGAAAACGGTGTGAAACTCGAAAATGGCGATTTCGTT  1182
Seq_2  1061  ------------------------------------------------------------  1060

Seq_1  1183  GAATTAGACATGCTGATATGTGCCACAGGTTTTGATGCCGTCGATGGCAACTATGTGCGC  1242
```

Where Seq_1 is the predicted pQR1034 sequence around the rplJ promoter and GenR, and Seq_2 is the sequencing result.

Appendix 2: CAI values for HEGs in *S. aureus* strains

The CAI values are arranged from the highest to lowest values for each strain. HEGs highlighted in grey correspond to genes that were also used in the CAI analysis of *S. carnosus*.

- *S. aureus* COL

GeneName	Function	CAI
-	hypothetical protein	0.905
-	hypothetical protein	0.854
-	hypothetical protein	0.853
hup	DNA-binding protein HU	0.851
-	PTS system, sorbitol-specific IIB component	0.844
rplA	50S ribosomal protein L1	0.837
eno	enolase	0.833
trxA	thioredoxin	0.831
atpE	ATP synthase subunit C	0.83
-	alkaline shock protein 23	0.827
gapA1	glyceraldehyde 3-phosphate dehydrogenase	0.826
isaA	immunodominant antigen A	0.823
rpsO	ribosomal protein S15	0.821
-	hypothetical protein	0.818

rplL	50S ribosomal protein L7/L12	0.817
-	hypothetical protein	0.817
rplU	ribosomal protein L21	0.813
rplM	50S ribosomal protein L13	0.813
tuf	elongation factor Tu	0.807
rpsD	30S ribosomal protein S4	0.804
rpmJ	ribosomal protein L36	0.802
-	cold shock protein, CSD family	0.798
ahpC	alkyl hydroperoxide reductase, C subunit	0.797
deoD2	purine nucleoside phosphorylase	0.797
glyS	glycyl-tRNA synthetase	0.795
fbaA	fructose-bisphosphate aldolase	0.794
rpsK	30S ribosomal protein S11	0.794
-	hypothetical protein	0.793
rplT	50S ribosomal protein L20	0.791
-	hypothetical protein	0.791
tig	trigger factor	0.789
-	peptidyl-prolyl cis-trans isomerase, cyclophilin-type	0.786
-	NADH dehydrogenase, putative	0.784
rpsT	ribosomal protein S20	0.783
clpP	ATP-dependent Clp protease, proteolytic subunit ClpP	0.783
pflB	formate acetyltransferase	0.783
rplY	ribosomal Protein L25	0.781
rplS	50S ribosomal protein L19	0.781
pdhD	dihydrolipoamide dehydrogenase	0.78
rpsR	ribosomal protein S18	0.78
tpiA	triosephosphate isomerase	0.777

rpsU	ribosomal protein S21	0.776
fusA	translation elongation factor G	0.776
lacD	tagatose 1,6-diphosphate aldolase	0.774
-	hypothetical protein	0.773
rpsI	ribosomal protein S9	0.773
agrD	accessory gene regulator protein D	0.772
rpsH	ribosomal protein S8	0.772
rpsP	ribosomal protein S16	0.77
-	LysM domain protein	0.769
tsf	elongation factor Ts	0.767
pgi	glucose-6-phosphate isomerase	0.767
-	hypothetical protein	0.766
rpsB	30S ribosomal protein S2	0.766
rplK	ribosomal protein L11	0.766
pdhC	pyruvate dehydrogenase complex E2 component, dihydrolipoamide acetyltransferase	0.766
glmS	D-fructose-6-phosphate amidotransferase	0.765
-	sceD protein, putative	0.764
pdhB	pyruvate dehydrogenase complex E1 component, beta subunit	0.764
sodA2	superoxide dismutase	0.762
rpoZ	DNA-directed RNA polymerase, omega subunit	0.762
rplQ	ribosomal protein L17	0.761
pyk	pyruvate kinase	0.761
rpmF	50S ribosomal protein L32	0.761
-	hypothetical protein	0.759
rplF	50S ribosomal protein L6	0.759
serS	seryl-tRNA synthetase	0.759
acpD	acyl carrier protein phosphodiesterase	0.758

rplJ	ribosomal protein L10	0.757
dltC	D-alanine--poly(phosphoribitol) ligase subunit 2	0.756
rpsN2	ribosomal protein S14	0.756
ilvC	ketol-acid reductoisomerase	0.755
-	staphyloxanthin biosynthesis protein	0.755
acpP	acyl carrier protein	0.755
rpmA	50S ribosomal protein L27	0.755
dnaK	dnaK protein	0.753
rpsE	ribosomal protein S5	0.753
-	cold shock protein, CSD family	0.753
rplX	50S ribosomal protein L24	0.753
efp	elongation factor P	0.753
fdaB	fructose-bisphosphate aldolase	0.752
-	HIT family protein	0.752
rpsG	30S ribosomal protein S7	0.75
-	cold shock protein, CSD family	0.749
-	hypothetical protein	0.749
pdhA	pyruvate dehydrogenase complex E1 component, alpha subunit	0.748
-	thioredoxin, putative	0.748
-	hypothetical protein	0.748
rplC	ribosomal protein L3	0.747
-	transferrin receptor	0.746
-	hypothetical protein	0.746
gnd	6-phosphogluconate dehydrogenase	0.745
atl	bifunctional autolysin	0.744
femC	glutamine synthetase FemC	0.744
greA	transcription elongation factor GreA	0.744

fmtB	fmtB protiein	0.744
-	hypothetical protein	0.744
rplE	ribosomal protein L5	0.744
ppaC	putative manganese-dependent inorganic pyrophosphatase	0.743
-	staphyloxanthin biosynthesis protein	0.742
-	fumarylacetoacetate hydrolase family protein	0.741
-	hypothetical protein	0.741
-	hypothetical protein	0.739
-	iron compound ABC transporter, iron compound-binding protein	0.738
-	hypothetical protein	0.738
-	copper ion binding protein	0.737
pgk	phosphoglycerate kinase	0.736
-	hypothetical protein	0.735
-	phosphate ABC transporter, phosphate-binding protein	0.735
rplR	ribosomal protein L18	0.734
-	hypothetical protein	0.734
guaB	inosine-5'-monophosphate dehydrogenase	0.733
fabG1	3-oxoacyl-(acyl-carrier-protein) reductase	0.733
-	hexulose-6-phosphate synthase, putative	0.732
lacB	galactose-6-phosphate isomerase	0.732
rplN	ribosomal protein L14	0.731
-	ferritins family protein	0.731
infB	translation initiation factor IF-2	0.731
rpmG2	ribosomal protein L33	0.731
icd	isocitrate dehydrogenase	0.731
-	IgG-binding protein SBI	0.73
-	hypothetical protein	0.73

rpsQ	30S ribosomal protein S17	0.73
purA	adenylosuccinate synthetase	0.728
-	ABC transporter, substrate-binding protein	0.728
mqo2	malate:quinone oxidoreductase	0.727
rpmH	ribosomal protein L34	0.727
rpsM	30S ribosomal protein S13	0.727
-	Dps family protein	0.726
-	nucleoside diphosphate kinase	0.726
sdhB	succinate dehydrogenase	0.726
sucD	succinyl-CoA synthetase alpha subunit	0.726
tal	transaldolase	0.725
groEL	chaperonin, 60 kDa	0.724
tyrS	tyrosyl-tRNA synthetase	0.724

- *S. aureus* Mu50

GeneName	Function	CAI
-	hypothetical protein	0.905
asp23	alkaline shock protein 23	0.864
-	similar to sigmaB-controlled protein	0.853
hu	DNA-binding protein II	0.852
-	hypothetical protein	0.85
rplA	50S ribosomal protein L1	0.837
eno	enolase	0.832
trxA	thioredoxin	0.832

atpE	ATP synthase subunit C	0.83
-	putative transcriptional regulator	0.822
gap	glyceraldehyde-3-phosphate dehydrogenase	0.82
rpsO	30Sribosomal protein S15	0.819
rplL	50S ribosomal protein L7/L12	0.818
rplU	50S ribosomal protein L21	0.813
rplM	50S ribosomal protein L13	0.813
isaA	immunodominant antigen A	0.812
rpsD	30S ribosomal protein S4	0.805
tig	trigger factor	0.804
tuf	elongation factor Tu	0.803
-	PTS system, galactitol-specific enzyme II B component	0.802
cspB	cold shock protein	0.799
rpmJ	50S ribosomal protein L36	0.797
ahpC	alkyl hydroperoxide reductase subunit C	0.797
fbaA	fructose-bisphosphate aldolase	0.795
rpsK	30S ribosomal protein S11	0.792
-	general stress protein-like protein	0.792
glyS	glycyl-tRNA synthetase	0.791
rplT	50S ribosomal protein L20	0.79
deoD	purine nucleoside phosphorylase	0.789
-	putative NADH dehydrogenase	0.784
rplY	50S ribosomal protein L25	0.784
rpsT	30S ribosomal protein S20	0.782
clpP	ATP-dependent Clp protease proteolytic subunit homolog	0.781
rplK	50S ribosomal protein L11	0.78
-	hypothetical protein	0.78

tpi	triosephosphate isomerase	0.778
pflB	formate acetyltransferase	0.778
rpsR	30S ribosomal protein S18	0.778
pdhD	dihydrolipoamide dehydrogenase	0.778
-	putative pit accessory protein	0.774
glmS	D-fructose-6-phosphate amidotransferase	0.774
rpsU	30S ribosomal protein S21	0.773
-	peptidyl-prolyl cis-trans isomerase homolog	0.772
fus	translational elongation factor G	0.77
rpsH	30S ribosomal protein S8	0.77
rpsI	30S ribosomal protein S9	0.769
rpsP	30S ribosomal protein S16	0.769
-	hypothetical protein	0.768
rpsB	30S ribosomal protein S2	0.768
-	secretory antigen SsaA homolog	0.767
rplJ	50S ribosomal protein L10	0.766
pykA	pyruvate kinase	0.766
-	hypothetical protein	0.764
-	hypothetical protein	0.764
tsf	elongation factor Ts	0.764
-	similar to SceD precursor	0.763
serS	seryl-tRNA synthetase	0.762
phdB	pyruvate dehydrogenase E1 component beta subunit	0.761
rplS	50S ribosomal protein L19	0.761
sodA	superoxide dismutase	0.76
rplQ	50S ribosomal protein L17	0.76
pgi	glucose-6-phosphate isomerase	0.758

-	probable DNA-directed RNA polymerase omega chain	0.758
rpsG	30S ribosomal protein S7	0.758
rpmF	50S ribosomal protein L32	0.757
rplF	50S ribosomal protein L6	0.756
hmrB	HmrB protein	0.755
rpsN	30S ribosomal protein S14	0.755
dnaK	DnaK protein	0.755
pdhC	dihydrolipoamide S-acetyltransferase component of pyruvate dehydrogenase complex E2	0.754
dltC	D-alanine--poly(phosphoribitol) ligase subunit 2	0.754
rpmA	50S ribosomal protein L27	0.753
-	elongation factor P	0.752
gnd	6-phosphogluconate dehydrogenase	0.752
rplX	50S ribosomal protein L24	0.752
rpsE	30S ribosomal protein S5	0.752
cspA	major cold shock protein	0.752
hit	Hit-like protein involved in cell-cycle regulation	0.75
lacD	tagatose 1,6-diphosphate aldolase	0.75
-	hypothetical protein	0.749
ssaA	secretory antigen precursor SsaA homolog	0.749
ilvC	ketol-acid reductoisomerase	0.748
glnA	glutamine-ammonia ligase	0.747
fmtB(mrp)	FmtB protein	0.746
-	lipoprotein	0.746
cspC	cold-shock protein C	0.746
-	5-oxo-1,2,5-tricarboxilic-3-penten acid decarboxylase	0.745
acpD	acyl carrier protein phosphodiesterase	0.745
pdhA	pyruvate dehydrogenase E1 component alpha subunit	0.744

rplE	50S ribosomal protein L5	0.744
-	phenylalanyl-tRNA synthetase homolog	0.742
greA	transcription elongation factor	0.742
csbD	sigmaB-controlled gene product	0.741
-	fructose-bisphosphate aldolase	0.741
dps	general stress protein 20U	0.739
-	partial autolysin	0.739
pgk	phosphoglycerate kinase	0.738
-	similar to secretory antigen precursor SsaA	0.737
-	similar to ferric hydroxamate receptor 1	0.737
-	hypothetical protein	0.736
-	hypothetical protein	0.736
-	putative manganese-dependent inorganic pyrophosphatase	0.735
-	similar to mercuric ion-binding protein	0.734
-	thioredoxine reductase	0.733
guaB	inositol-monophosphate dehydrogenase	0.732
fabG	3-oxoacyl-(acyl-carrier protein) reductase	0.732
lacG	6-phospho-beta-galactosidase	0.731
rplN	50S ribosomal protein L14	0.731
infB	translation initiation factor IF-2	0.731
-	hypothetical protein	0.731
-	hypothetical protein	0.73
rplR	50S ribosomal protein L18	0.73
lacB	galactose-6-phosphate isomerase	0.729
-	putative hexulose-6-phosphate synthase	0.729
citC	isocitrate dehydrogenase	0.729
sbi	IgG-binding protein	0.729

-	ferritin	0.728
rpmG	50S ribosomal protein L33	0.728
rpsM	30S ribosomal protein S13	0.728
groEL	GroEL protein	0.728
pyrE	orotate phosphoribosyltransferase	0.727
rpsL	30S ribosomal protein S12	0.726
rpsC	30S ribosomal protein S3	0.725
rplI	50S ribosomal protein L9	0.723
-	similar to ABC transporter substrate-binding protein	0.723
purA	adenylosuccinate synthase	0.723
sdhB	succinate dehydrogenase	0.723
-	hypothetical protein	0.723
rpmH	50S ribosomal protein L34	0.722
-	hypothetical protein	0.722
-	similar to thioredoxin	0.721
tyrS	tyrosyl-tRNA synthetase	0.721
mqo2	malate:quinone oxidoreductase	0.72
-	alpha-acetolactate decarboxylase	0.718

Appendix 3: DoE design for optimization of the staphylococcal electroporation protocol

	A	B	C	D	E	F
1	hypertonic	21 kv/cm 1.1	B2-	0.1	-	Sucrose
2	water	21 kv/cm 1.1	B2-	0.1	-	no sucrose
3	hypertonic	2.5 kv/cm 2.5	B2-	0.1	-	no sucrose
4	water	2.5 kv/cm 2.5	B2-	0.1	-	Sucrose
5	hypertonic	21 kv/cm 1.1	B2	0.1	-	no sucrose
6	water	21 kv/cm 1.1	B2	0.1	-	Sucrose
7	hypertonic	2.5 kv/cm 2.5	B2	0.1	-	Sucrose
8	water	2.5 kv/cm 2.5	B2	0.1	-	no sucrose
9	hypertonic	21 kv/cm 1.1	B2-	0.5	-	no sucrose
10	water	21 kv/cm 1.1	B2-	0.5	-	Sucrose
11	hypertonic	2.5 kv/cm 2.5	B2-	0.5	-	Sucrose
12	water	2.5 kv/cm 2.5	B2-	0.5	-	no sucrose
13	hypertonic	21 kv/cm 1.1	B2	0.5	-	Sucrose
14	water	21 kv/cm 1.1	B2	0.5	-	no sucrose
15	hypertonic	2.5 kv/cm 2.5	B2	0.5		no sucrose
16	water	2.5 kv/cm 2.5	B2	0.5	-	Sucrose
17	hypertonic	21 kv/cm 1.1	B2-	0.1	+	Sucrose
18	water	21 kv/cm 1.1	B2-	0.1	+	no sucrose
19	hypertonic	2.5 kv/cm 2.5	B2-	0.1	+	no sucrose
20	water	2.5 kv/cm 2.5	B2-	0.1	+	Sucrose
21	hypertonic	21 kv/cm 1.1	B2	0.1	+	no sucrose
22	water	21 kv/cm 1.1	B2	0.1	+	Sucrose
23	hypertonic	2.5 kv/cm 2.5	B2	0.1	+	Sucrose
24	water	2.5 kv/cm 2.5	B2	0.1	+	no sucrose
25	hypertonic	21 kv/cm 1.1	B2-	0.5	+	no sucrose
26	water	21 kv/cm 1.1	B2-	0.5	+	Sucrose
27	hypertonic	2.5 kv/cm 2.5	B2-	0.5	+	Sucrose
28	water	2.5 kv/cm 2.5	B2-	0.5	+	no sucrose
29	hypertonic	21 kv/cm 1.1	B2	0.5	+	Sucrose
30	water	21 kv/cm 1.1	B2	0.5	+	no sucrose
31	hypertonic	2.5 kv/cm 2.5	B2	0.5	+	no sucrose
32	water	2.5 kv/cm 2.5	B2	0.5	+	Sucrose